Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 1   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Effect of Internet-Based Rehabilitation Programs on Improvement of Pain and Physical Function in Patients with Knee Osteoarthritis: Systematic Review and Meta-analysis of Randomized Controlled Trials (e21542)
Su-Hang Xie, Qian Wang, Li-Qiong Wang, Lin Wang, Kang-Ping Song, Cheng-Qi He. 9

Assessment of Psychological Distress in Adults With Type 2 Diabetes Mellitus Through Technologies: Literature Review (e17740)
Giulia Bassi, Silvia Gabrielli, Valeria Donisi, Sara Carbone, Stefano Forti, Silvia Salcuni. 21

Effect of Interactive eHealth Interventions on Improving Medication Adherence in Adults With Long-Term Medication: Systematic Review (e18901)
Bart Pouls, Johanna Vriezekolk, Charlotte Bekker, Annemiek Linn, Hein van Onzenoort, Marcia Vervloet, Sandra van Dulmen, Bart van den Bemt. 34

Smartphone-Based Interventions and Internalizing Disorders in Youth: Systematic Review and Meta-analysis (e16490)
Adrian Buttazzoni, Keshbir Brar, Leia Minaker. 50

How the Use of a Patient-Accessible Health Record Contributes to Patient-Centered Care: Scoping Review (e17655)
Janine Benjamins, Annemien Haveman-Nies, Marian Gunnink, Annemieke Goudkuil, Emely de Vet. 62

Young People's Use of Digital Health Technologies in the Global North: Narrative Review (e18286)
Deborah Lupton. 76

The Role of Health Technologies in Multicomponent Primary Care Interventions: Systematic Review (e20195)
Geronimo Jimenez, David Matchar, Choon Koh, Rianne van der Kleij, Niels Chavannes, Josip Car. 88

Applications and Recruitment Performance of Web-Based Respondent-Driven Sampling: Scoping Review (e17564)
Yannick Helms, Nora Hamdiui, Mirjam Kretzschmar, Luis Rocha, Jim van Steenbergen, Linus Bengtsson, Anna Thorson, Aura Timen, Mart Stein. 105

Implementation of Telehealth Services to Assess, Monitor, and Treat Neurodevelopmental Disorders: Systematic Review (e22619)
Althea Valentine, Sophie Hall, Emma Young, Beverley Brown, Madeleine Groom, Chris Hollis, Charlotte Hall. 117

Factors Affecting the Implementation, Use, and Adoption of Real-Time Location System Technology for Persons Living With Cognitive Disabilities in Long-term Care Homes: Systematic Review (e22831)
Alisa Grigorovich, Yalinie Kulandaivelu, Kristine Newman, Andria Bianchi, Shehroz Khan, Andrea Iaboni, Josephine McMurray. 131

Determining the Intellectual Structure and Academic Trends of Smart Home Health Care Research: Coword and Topic Analyses (e19625)
Hyo-Jin Kang, Jieun Han, Gyu Kwon. 144

Understanding eHealth Cognitive Behavioral Therapy Targeting Substance Use: Realist Review (e20557)
Farhud Shams, James Wong, Mohammadali Nikoo, Ava Outadi, Ehsan Moazen-Zadeh, Mostafa Kamel, Michael Song, Kerry Jang, Reinhard Krausz. 160

Possibilities, Problems, and Perspectives of Data Collection by Mobile Apps in Longitudinal Epidemiological Studies: Scoping Review (e17691)
Florian Fischer, Sina Kleen. 170

Data Quality and Cost-effectiveness Analyses of Electronic and Paper-Based Interviewer-Administered Public Health Surveys: Systematic Review (e21382)
Atinkut Zeleke, Tolga Naziyok, Fleur Fritz, Lara Christianson, Rainer Röhrig. 183

Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for Processing Free Text in Health Care: Systematic Scoping Review (e24594)
Christophe Gaudet-Blavignac, Vasiliki Foufi, Mina Bjelogrlic, Christian Lovis. 207

Prevalence of Health Misinformation on Social Media: Systematic Review (e17187)
Victor Suarez-Lledo, Javier Alvarez-Galvez. 450

Spread, Scale-up, and Sustainability of Video Consulting in Health Care: Systematic Review and Synthesis Guided by the NASSS Framework (e23775)
Hannah James, Chrysanthi Papoutsi, Joseph Wherton, Trisha Greenhalgh, Sara Shaw. 487

Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review (e17828)
Alaa Abd-Alrazaq, Mohannad Alajlani, Nashva Ali, Kerstin Denecke, Bridgette Bewick, Mowafa Househ. 888

Viewpoints

Building Health Services in a Rapidly Changing Landscape: Lessons in Adaptive Leadership and Pivots in a COVID-19 Remote Monitoring Program (e25507)
Celia Laur, Payal Agarwal, Geetha Mukerji, Elaine Goulbourne, Hayley Baranek, Laura Pus, R Bhatia, Danielle Martin, Onil Bhattacharyya. . .
2 2 5

Going From Zero to 100 in Remote Dementia Research: A Practical Guide (e24098)
Megan O'Connell, Shirin Vellani, Sheryl Robertson, Hannah O'Rourke, Kathy McGilton. 233

Patient Care During the COVID-19 Pandemic: Use of Virtual Care (e20621)
Andy Wong, Rashaad Bhyat, Siddhartha Srivastava, Lysa Boissé Lomax, Ramana Appireddy. 1115

Digital Health Tools for Managing Noncommunicable Diseases During and After the COVID-19 Pandemic: Perspectives of Patients and Caregivers (e25652)
Alessandro Monaco, Katie Palmer, Nicolaj Holm Ravn Faber, Irene Kohler, Mitchell Silva, Anita Vatland, Joop van Griensven, Mariano Votta, Donna Walsh, Vincent Clay, Mehmet Yazicioglu, Danute Ducinskiene, Shaantanu Donde. 1202

Original Papers

Digital Health Needs of Women With Postpartum Depression: Focus Group Study (e18934)
Madison Lackie, Julia Parrilla, Brynn Lavery, Andrea Kennedy, Deirdre Ryan, Barbara Shulman, Lori Brotto. . . . 241

Uptake of Proactively Offered Online and Telephone Support Services Targeting Multiple Health Risk Behaviors Among Vocational Education Students: Process Evaluation of a Cluster Randomized Controlled Trial (e19737)
Prince Atorkey, Christine Paul, Billie Bonevski, John Wiggers, Aimee Mitchell, Emma Byrnes, Christophe Lecathelinais, Flora Tzelepis. . . . 259

Resource-Based Internet Intervention (Med-Stress) to Improve Well-Being Among Medical Professionals: Randomized Controlled Trial (e21445)
Ewelina Smoktunowicz, Magdalena Lesnierowska, Per Carlbring, Gerhard Andersson, Roman Cieslak. . . . 276

Effects of Website Interactivity on Skin Cancer–Related Intentions and User Experience: Factorial Randomized Experiment (e18299)
Zhaomeng Niu, Jessica Willoughby, Elliot Coups, Jerod Stapleton. . . . 294

Effects of Web-Based Social Connectedness on Older Adults' Depressive Symptoms: A Two-Wave Cross-Lagged Panel Study (e21275)
Juwon Hwang, Catalina Toma, Junhan Chen, Dhavan Shah, David Gustafson, Marie-Louise Mares. . . . 308

Digital Intervention With Lifestyle Coach Support to Target Dietary and Physical Activity Behaviors of Adults With Nonalcoholic Fatty Liver Disease: Systematic Development Process of VITALISE Using Intervention Mapping (e20491)
Kate Hallsworth, Stuart McPherson, Quentin Anstee, Darren Flynn, Laura Haigh, Leah Avery. . . . 321

Efficacy of a Transdiagnostic Self-Help Internet Intervention for Reducing Depression, Anxiety, and Suicidal Ideation in Adults: Randomized Controlled Trial (e22698)
Philip Batterham, Alison Calear, Louise Farrer, Amelia Gulliver, Ella Kurz. . . . 336

Effectiveness of Smartphone-Based Mindfulness Training on Maternal Perinatal Depression: Randomized Controlled Trial (e23410)
Yaoyao Sun, Yanyan Li, Juan Wang, Qingyi Chen, Alessandra Bazzano, Fenglin Cao. . . . 349

General Behavioral Engagement and Changes in Clinical and Cognitive Outcomes of Patients with Type 2 Diabetes Using the Time2Focus Mobile App for Diabetes Education: Pilot Evaluation (e17537)
Bryan Batch, Susan Spratt, Dan Blalock, Chad Benditz, Andi Weiss, Rowena Dolor, Alex Cho. . . . 372

A Mobile App to Support Clinical Diagnosis of Upper Respiratory Problems (eHealthResp): Co-Design Approach (e19194)
João Moura, Ana Almeida, Fátima Roque, Adolfo Figueiras, Maria Herdeiro. . . . 382

The Role of Technology and Social Media Use in Sleep-Onset Difficulties Among Italian Adolescents: Cross-sectional Study (e20319)
Nirosha Varghese, Eugenio Santoro, Alessandra Lugo, Juan Madrid-Valero, Simone Ghislandi, Aleksandra Torbica, Silvano Gallus. . . . 397

Patterns of Routes of Administration and Drug Tampering for Nonmedical Opioid Consumption: Data Mining and Content Analysis of Reddit Discussions (e21212)
Duilio Balsamo, Paolo Bajardi, Alberto Salomone, Rossano Schifanella. . . . 410

“Down the Rabbit Hole” of Vaccine Misinformation on YouTube: Network Exposure Study (e23262)
Lu Tang, Kayo Fujimoto, Muhammad Amith, Rachel Cunningham, Rebecca Costantini, Felicia York, Grace Xiong, Julie Boom, Cui Tao. . . . 427

A Novel Machine Learning Framework for Comparison of Viral COVID-19–Related Sina Weibo and Twitter Posts: Workflow Development and Content Analysis (e24889)

Shi Chen, Lina Zhou, Yunya Song, Qian Xu, Ping Wang, Kanlun Wang, Yaorong Ge, Daniel Janies. 436

Perceived Factors Influencing the Public Intention to Use E-Consultation: Analysis of Web-Based Survey Data (e21834)

Miaojie Qi, Jiyu Cui, Xing Li, Youli Han. 467

Adherence of Mobile App-Based Surveys and Comparison With Traditional Surveys: eCohort Study (e24773)

Chathurangi Pathiravasan, Yuankai Zhang, Ludovic Trinquart, Emelia Benjamin, Belinda Borrelli, David McManus, Vik Kheterpal, Honghuang Lin, Mayank Sardana, Michael Hammond, Nicole Spartano, Amy Dunn, Eric Schramm, Christopher Nowak, Emily Manders, Hongshan Liu, Jelena Kornej, Chunyu Liu, Joanne Murabito. 502

Relationship Between Internet Use and Change in Health Status: Panel Study of Young Adults (e22051)

Amanda Hunsaker, Eszter Hargittai, Marina Micheli. 516

Use of eHealth by Patients With Rheumatoid Arthritis: Observational, Cross-sectional, Multicenter Study (e19998)

Marion Magnol, Berard Eleonore, Rempenault Claire, Benjamin Castagne, Marine Pugibet, Cédric Lukas, Anne Tournadre, Pascale Vergne-Salle, Thomas Barnetche, Marie-Elise Truchetet, Adeline Ruyssen-Witrand. 528

Content-Based Recommender Support System for Counselors in a Suicide Prevention Chat Helpline: Design and Evaluation Study (e21690)

Salim Salmi, Saskia Mérelle, Renske Gilissen, Willem-Paul Brinkman. 535

Gaining Insights Into the Estimation of the Circadian Rhythms of Social Activity in Older Adults From Their Telephone Call Activity With Statistical Learning: Observational Study (e22339)

Timothée Aubourg, Jacques Demongeot, Nicolas Vuillerme. 544

Sharing Patient-Controlled Real-World Data Through the Application of the Theory of Commons: Action Research Case Study (e16842)

Andreas Hager, Staffan Lindblad, Mats Brommels, Stina Salomonsson, Carolina Wannheden. 555

Innovation in Pain Rehabilitation Using Co-Design Methods During the Development of a Relapse Prevention Intervention: Case Study (e18462)

Stefan Elbers, Christa van Gessel, Reint Renes, Remko van der Lugt, Harriët Wittink, Sander Hermsen. 567

Ride-Hailing Services and Alcohol Consumption: Longitudinal Analysis (e15402)

Gordon Burtch, Brad Greenwood, Jeffrey McCullough. 577

Patients' Perspectives About Factors Affecting Their Use of Electronic Personal Health Records in England: Qualitative Analysis (e17500)

Alaa Abd-Alrazaq, Zeineb Safi, Bridgette Bewick, Mowafa Househ, Peter Gardner. 586

The Causal Relationship Between Portal Usage and Self-Efficacious Health Information–Seeking Behaviors: Secondary Analysis of the Health Information National Trends Survey Data (e17782)

Jaeyoung Park, Muxuan Liang, Jordan Alpert, Richard Brown, Xiang Zhong. 599

Application of an Adaptive, Digital, Game-Based Approach for Cognitive Assessment in Multiple Sclerosis: Observational Study (e24356)

Wan-Yu Hsu, William Rowles, Joaquin Anguera, Chao Zhao, Annika Anderson, Amber Alexander, Simone Sacco, Roland Henry, Adam Gazzaley, Riley Bove. 608

Development and Implementation of a Decision Support System to Improve Control of Hypertension and Diabetes in a Resource-Constrained Area in Brazil: Mixed Methods Study (e18872)

Milena Marcolino, João Oliveira, Christiane Cimini, Junia Maia, Vânia Pinto, Thábata Sá, Kaique Amancio, Lissandra Coelho, Leonardo Ribeiro, Clareci Cardoso, Antonio Ribeiro. 623

Use of BERT (Bidirectional Encoder Representations from Transformers)-Based Deep Learning Method for Extracting Evidences in Chinese Radiology Reports: Development of a Computer-Aided Liver Cancer Diagnosis Framework (e19689)

Honglei Liu, Zhiqiang Zhang, Yan Xu, Ni Wang, Yanqun Huang, Zhenghan Yang, Rui Jiang, Hui Chen. 641

Disease Concept-Embedding Based on the Self-Supervised Method for Medical Information Extraction from Electronic Health Records and Disease Retrieval: Algorithm Development and Validation Study (e25113)

Yen-Pin Chen, Yuan-Hsun Lo, Feipei Lai, Chien-Hua Huang. 652

A Newly Developed Web-Based Resource on Genetic Eye Disorders for Users With Visual Impairment (Gene.Vision): Usability Study (e19151)

Jian Yeong, Peter Thomas, James Buller, Mariya Moosajee. 666

COVID-19 Misinformation Trends in Australia: Prospective Longitudinal National Survey (e23805)

Kristen Pickles, Erin Cvejic, Brooke Nickel, Tessa Copp, Carissa Bonner, Julie Leask, Julie Ayre, Carys Batcup, Samuel Cornell, Thomas Dakin, Rachael Dodd, Jennifer Isautier, Kirsten McCaffery. 683

Evaluation of the Liver Disease Information in Baidu Encyclopedia and Wikipedia: Longitudinal Study (e17680)

Fei Sun, Fuchun Yang, Shusen Zheng. 697

Adherence of Internet-Based Cancer Risk Assessment Tools to Best Practices in Risk Communication: Content Analysis (e23318)

Erika Waters, Jeremy Foust, Laura Scherer, Amy McQueen, Jennifer Taber. 705

Patient Perspectives on Health Data Privacy and Implications for Adverse Drug Event Documentation and Communication: Qualitative Study (e21452)

Serena Small, Corinne Hohl, Ellen Balka. 718

Improving Uptake of a National Web-Based Psychoeducational Workshop for Informal Caregivers of Veterans: Mixed Methods Implementation Evaluation (e16495)

Veronica Yank, Randall Gale, Andrea Nevedal, Leonore Okwara, Christopher Koenig, Ranak Trivedi, Nancy Dupke, Margaret Kabat, Steven Asch. 728

Lifelog Data-Based Prediction Model of Digital Health Care App Customer Churn: Retrospective Observational Study (e22184)

Hongwook Kwon, Ho Kim, Jaeil An, Jae-Ho Lee, Yu Park. 748

Factors Associated With Dropout of Participants in an App-Based Child Injury Prevention Study: Secondary Data Analysis of a Cluster Randomized Controlled Trial (e21636)

Jie Li, Peishan Ning, Peixia Cheng, David Schwebel, Yang Yang, Xiang Wei, Jieyi He, Wanhui Wang, Ruotong Li, Guoqing Hu. 760

Virtual Reality Relaxation for Patients With a Psychiatric Disorder: Crossover Randomized Controlled Trial (e17233)

Wim Veling, Bart Lestestuiver, Marieke Jongma, H Hoenders, Catheleine van Driel. 770

Virtual Reality Cognitive Training Among Individuals With Alcohol Use Disorder Undergoing Residential Treatment: Pilot Randomized Controlled Trial (e18482)

Pedro Gamito, Jorge Oliveira, Marcelo Matias, Elsa Cunha, Rodrigo Brito, Paulo Lopes, Alberto Deus. 784

Perceived Needs Versus Predisposing/Enabling Characteristics in Relation to Internet Cancer Information Seeking Among the US and Chinese Public: Comparative Survey Research (e24733)
Di Zhang, Hongchao Hu, Zhen Shi, Biao Li. 795

A Bayesian Network–Based Browsing Model for Patients Seeking Radiology-Related Information on Hospital Websites: Development and Usability Study (e14794)
Ryusuke Suzuki, Teppei Suzuki, Shintaro Tsuji, Kensuke Fujiwara, Hiroko Yamashina, Akira Endoh, Katsuhiko Ogasawara. 810

Effects of Objective and Subjective Health Literacy on Patients' Accurate Judgment of Health Information and Decision-Making Ability: Survey Study (e20457)
Peter Schulz, Annalisa Pessina, Uwe Hartung, Serena Petrocchi. 828

Role of Health Literacy in Health-Related Information-Seeking Behavior Online: Cross-sectional Study (e14088)
Hee Lee, Seok Jin, Carrie Henning-Smith, Jongwook Lee, Jaegoo Lee. 841

Differences in Mode Preferences, Response Rates, and Mode Effect Between Automated Email and Phone Survey Systems for Patients of Primary Care Practices: Cross-Sectional Study (e21240)
Sharon Johnston, William Hogg, Sabrina Wong, Fred Burge, Sandra Peterson. 851

Risk Stratification for Early Detection of Diabetes and Hypertension in Resource-Limited Settings: Machine Learning Analysis (e20123)
Justin Boutilier, Timothy Chan, Manish Ranjan, Sarang Deo. 861

Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study (e19928)
Xiangmin Fan, Daren Chao, Zhan Zhang, Dakuo Wang, Xiaohua Li, Feng Tian. 872

Deliberative and Paternalistic Interaction Styles for Conversational Agents in Digital Health: Procedure and Validation Through a Web-Based Experiment (e22919)
Theresa Schachner, Christoph Gross, Andrea Hasl, Florian v Wangenheim, Tobias Kowatsch. 903

Economic Burden and Health Care Access for Patients With Inflammatory Bowel Diseases in China: Web-Based Survey Study (e20629)
Qiao Yu, Chunpeng Zhu, Shuyi Feng, Liyi Xu, Shurong Hu, Hao Chen, Hanwen Chen, Sheng Yao, Xiaoying Wang, Yan Chen. 916

Retracted: "Assessing the Dissemination of COVID-19 Articles Across Social Media With Altmetric and PlumX Metrics: Correlational Study" (e21408)
Haley Tornberg, Carine Moezinia, Chapman Wei, Simone Bernstein, Chaplin Wei, Refka Al-Beyati, Theodore Quan, David Diemert. 930

The HealthChain Blockchain for Electronic Health Records: Development Study (e13556)
Yonggang Xiao, Bin Xu, Wenhao Jiang, Yunjun Wu. 938

Rapid COVID-19 Screening Based on Self-Reported Symptoms: Psychometric Assessment and Validation of the EPICOVID19 Short Diagnostic Scale (e23897)
Luca Bastiani, Loredana Fortunato, Stefania Pieroni, Fabrizio Bianchi, Fulvio Adorni, Federica Prinelli, Andrea Giacomelli, Gabriele Pagani, Stefania Maggi, Caterina Trevisan, Marianna Noale, Nithiya Jesuthasan, Aleksandra Sojic, Carla Pettenati, Massimo Andreoni, Raffaele Antonelli Incalzi, Massimo Galli, Sabrina Molinaro. 957

Accurately Differentiating Between Patients With COVID-19, Patients With Other Viral Infections, and Healthy Individuals: Multimodal Late Fusion Learning Approach (e25535)
Ming Xu, Liu Ouyang, Lei Han, Kai Sun, Tingting Yu, Qian Li, Hua Tian, Lida Safarnejad, Hengdong Zhang, Yue Gao, Forrest Bao, Yuanfang Chen, Patrick Robinson, Yaorong Ge, Baoli Zhu, Jie Liu, Shi Chen. 969

Investigating and Improving the Accuracy of US Citizens' Beliefs About the COVID-19 Pandemic: Longitudinal Survey Study (e24069)
Aart van Stekelenburg, Gabi Schaap, Harm Veling, Moniek Buijzen. . . . 986

Digital Health Literacy and Web-Based Information-Seeking Behaviors of University Students in Germany During the COVID-19 Pandemic: Cross-sectional Survey Study (e24097)
Kevin Dadaczynski, Orkan Okan, Melanie Messer, Angela Leung, Rafaela Rosário, Emily Darlington, Katharina Rathmann. . . . 1001

Asynchronous Distance Learning of the National Institutes of Health Stroke Scale During the COVID-19 Pandemic (E-Learning vs Video): Randomized Controlled Trial (e23594)
Mélanie Suppan, Loric Stuby, Emmanuel Carrera, Philippe Cottet, Avinash Koka, Frédéric Assal, Georges Savoldelli, Laurent Suppan. . . . 1018

SARS-CoV-2 Surveillance in the Middle East and North Africa: Longitudinal Trend Analysis (e25830)
Lori Post, Emily Marogi, Charles Moss, Robert Murphy, Michael Ison, Chad Achenbach, Danielle Resnick, Lauren Singh, Janine White, Michael Boctor, Sarah Welch, James Oehmke. . . . 1037

Unfolding the Determinants of COVID-19 Vaccine Acceptance in China (e26089)
Fulian Yin, Zhaoliang Wu, Xinyu Xia, Meiqi Ji, Yanyan Wang, Zhiwen Hu. . . . 1048

Maternal Mental Health Status and Approaches for Accessing Antenatal Care Information During the COVID-19 Epidemic in China: Cross-Sectional Study (e18722)
Hong Jiang, Longmei Jin, Xu Qian, Xu Xiong, Xuena La, Weiyi Chen, Xiaoguang Yang, Fengyun Yang, Xinwen Zhang, Nazhakaiti Abudukelimu, Xingying Li, Zhenyu Xie, Xiaoling Zhu, Xiaohua Zhang, Lifeng Zhang, Li Wang, Lingling Li, Mu Li. . . . 1065

Mental Health Among Medical Professionals During the COVID-19 Pandemic in Eight European Countries: Cross-sectional Survey Study (e24983)
Svenja Hummel, Neele Oetjen, Junfeng Du, Elisabetta Posenato, Rosa Resende de Almeida, Raquel Losada, Oscar Ribeiro, Vincenza Frisardi, Louise Hopper, Asarnusch Rashid, Habib Nasser, Alexandra König, Gottfried Rudofsky, Steffi Weidt, Ali Zafar, Nadine Gronewold, Gwendolyn Mayer, Jobst-Hendrik Schultz. . . . 1079

Assessment of Internet Hospitals in China During the COVID-19 Pandemic: National Cross-Sectional Data Analysis Study (e21825)
Xingyan Xu, Yingying Cai, Siying Wu, Jianhui Guo, Le Yang, Jieli Lan, Yi Sun, Bingbing Wang, Jieyu Wu, Tinggui Wang, Shuna Huang, Yawen Lin, Yuduan Hu, Mingjun Chen, Xuecai Gao, Xiaoxu Xie. . . . 1091

Use of a Remote Oncology Pharmacy Service Platform for Patients With Cancer During the COVID-19 Pandemic: Implementation and User Acceptance Evaluation (e24619)
Zhuo-Jia Chen, Wei-Ting Liang, Qing Liu, Rong He, Qian-Chao Chen, Qiu-Feng Li, Yao Zhang, Xiao-Dong Du, Ying Pan, Shu Liu, Xiao-Yan Li, Xue Wei, He Huang, Hong-Bing Huang, Tao Liu. . . . 1103

Social, Cognitive, and eHealth Mechanisms of COVID-19–Related Lockdown and Mandatory Quarantine That Potentially Affect the Mental Health of Pregnant Women in China: Cross-Sectional Survey Study (e24495)
Xue Yang, Bo Song, Anise Wu, Phoenix Mo, Jiangli Di, Qian Wang, Joseph Lau, Linhong Wang. . . . 1124

Socioeconomic Disparities in Social Distancing During the COVID-19 Pandemic in the United States: Observational Study (e24591)
Romain Garnier, Jan Benetka, John Kraemer, Shweta Bansal. . . . 1141

Toward Using Twitter for Tracking COVID-19: A Natural Language Processing Pipeline and Exploratory Data Set (e25314)
Ari Klein, Arjun Magge, Karen O'Connor, Jesus Flores Amaro, Davy Weissenbacher, Graciela Gonzalez Hernandez. . . . 1149

Adequacy of Web-Based Activities as a Substitute for In-Person Activities for Older Persons During the COVID-19 Pandemic: Survey Study (e25848)
Jiska Cohen-Mansfield, Aline Muff, Guy Meschiany, Shahar Lev-Ari. . . . 1155

Expanding Telemonitoring in a Virtual World: A Case Study of the Expansion of a Heart Failure Telemonitoring Program During the COVID-19 Pandemic (e26165)

Sahr Wali, Milena Guessi Margarido, Amika Shah, Patrick Ware, Michael McDonald, Mary O'Sullivan, Juan Duero Posada, Heather Ross, Emily Seto. 1168

Impact of COVID-19 on Physical Activity Among 10,000 Steps Members and Engagement With the Program in Australia: Prospective Study (e23946)

Quyen To, Mitch Duncan, Anetta Van Itallie, Corneel Vandelanotte. 1180

Geographic Distribution of Mental Health Problems Among Chinese College Students During the COVID-19 Pandemic: Nationwide, Web-Based Survey Study (e23126)

Xiaoyan Wu, Shuman Tao, Yi Zhang, Shiyue Li, Le Ma, Yizhen Yu, Guilong Sun, Tingting Li, Fangbiao Tao. 1189

Application of In-Home Monitoring Data to Transition Decisions in Continuing Care Retirement Communities: Usability Study (e18806)

Katherine Wild, Nicole Sharma, Nora Mattek, Jason Karlawish, Thomas Riley, Jeffrey Kaye. 1211

Corrigenda

Correction: A Newly Developed Web-Based Resource on Genetic Eye Disorders for Users With Visual Impairment (Gene.Vision): Usability Study (e27330)

Jian Yeong, Peter Thomas, James Buller, Mariya Moosajee. 951

Corrigenda and Addendas

Correction: Application of an Adaptive, Digital, Game-Based Approach for Cognitive Assessment in Multiple Sclerosis: Observational Study (e27440)

Wan-Yu Hsu, William Rowles, Joaquin Anguera, Chao Zhao, Annika Anderson, Amber Alexander, Simone Sacco, Roland Henry, Adam Gazzaley, Riley Bove. 953

Correction: Electronic Cognitive Screen Technology for Screening Older Adults With Dementia and Mild Cognitive Impairment in a Community Setting: Development and Validation Study (e26724)

Joyce Chan, Adrian Wong, Brian Yiu, Hazel Mok, Patti Lam, Pauline Kwan, Amany Chan, Vincent Mok, Kelvin Tsoi, Timothy Kwok. 955

Review

Effect of Internet-Based Rehabilitation Programs on Improvement of Pain and Physical Function in Patients with Knee Osteoarthritis: Systematic Review and Meta-analysis of Randomized Controlled Trials

Su-Hang Xie[1,2,3*], PT, MM; Qian Wang[1,2,3*], MD, PhD; Li-Qiong Wang[1,2,3], PT, PhD; Lin Wang[1,2,3], PT, MM; Kang-Ping Song[1,2,3], PT, MM; Cheng-Qi He[2,3], MD, PhD

[1]School of Rehabilitation Sciences, West China School of Medicine, Sichuan University, Chengdu, Sichuan, China
[2]Rehabilitation Medicine Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China
[3]Key Laboratory of Rehabilitation Medicine in Sichuan Province, Chengdu, Sichuan, China
[*]these authors contributed equally

Corresponding Author:
Cheng-Qi He, MD, PhD
Rehabilitation Medicine Center
West China Hospital
Sichuan University
No 37 Guoxue Street
Wuhou District
Chengdu, Sichuan, 610041
China
Phone: 86 28 8542 2847
Fax: 86 28 8542 3819
Email: hxkfhcq2015@126.com

Abstract

Background: Osteoarthritis (OA) is a chronic, debilitating, and degenerative joint disease. However, it is difficult for patients with knee OA to access conventional rehabilitation when discharging from the hospital. Internet-based rehabilitation is one of the promising telemedicine strategies to provide a means combining monitoring, guidance, and treatment for patients with knee OA.

Objective: The aim of this study was to conduct a systematic review and meta-analysis for assessing the effect of internet-based rehabilitation programs on pain and physical function in patients with knee OA.

Methods: Keywords related to knee OA and internet-based rehabilitation were systematically searched in the Web of Science, MEDLINE, EMBASE, CENTRAL, Scopus, PEDro (Physiotherapy Evidence Database), CNKI, SinoMed, and WANFANG databases from January 2000 to April 2020. Only randomized controlled trials were included. The authors independently screened the literature. The main outcome measures were focused on pain and physical function. A meta-analysis was performed on the collected data. Review Manager (RevMan, version 5.3) was used for all analyses.

Results: The systematic review identified 6 randomized controlled trials, 4 of which were included in the meta-analysis, comprising a total of 791 patients with knee OA. The meta-analysis with the fixed-effects model showed that the internet-based rehabilitation programs could significantly alleviate the osteoarthritic pain for patients compared with conventional rehabilitation (standardized mean difference [SMD] –0.21, 95% CI −0.4 to –0.01, *P*=.04). No significant difference was found in the improvement of physical function in patients with knee OA compared with conventional rehabilitation within 2 to 12 months (SMD –0.08, 95% CI −0.27 to 0.12, *P*=.43).

Conclusions: This systematic review shows that internet-based rehabilitation programs could improve the pain but not physical function for patients with knee OA. However, there was a very small number of studies that could be included in the review and meta-analysis. Thus, further studies with large sample sizes are warranted to promote the effectiveness of internet-based rehabilitation and to develop its personalized design.

(J Med Internet Res 2021;23(1):e21542) doi:10.2196/21542

KEYWORDS

internet-based rehabilitation; knee; osteoarthritis; pain; physical function; meta-analysis; review; telerehabilitation; eHealth; telemedicine

Introduction

Osteoarthritis (OA) is a chronic, debilitating, and degenerative joint disease, which is widely considered as a significant threat to healthy aging [1-3]. A recent estimation revealed that approximately 250 million individuals are suffering from OA worldwide, and the knee joint is the most frequently affected joint with an incidence of 16%-17% among people aged 50 to 75 years [4,5]. Chronic pain and impaired physical function are recognized as the main issues affecting quality of life for patients with knee OA [6,7]. Physical therapy is one of the effective methods that is commonly prescribed for patients with knee OA to alleviate pain and improve physical function [8-10]. However, it is difficult for patients with knee OA, who are mainly middle-aged to elderly, to access conventional rehabilitation programs that are monitored and guided by physicians or therapists when discharging from the hospital [4]. Thus, it is necessary to develop telerehabilitation strategies to provide the opportunity to access professional rehabilitation programs and guides for improving the long-term outcomes of pain and physical function for patients with knee OA.

The advent of telemedicine has facilitated the access of patients to real-time communication with professional physicians or therapists [11-13]. Telerehabilitation has been developed in the fields of physical medicine and rehabilitation to support continuous rehabilitation services for patients with disabilities [14]. Several types of telerehabilitation services have been proposed, including video conference, telephone conference, and web-based knowledge platforms [15-17]. Of these, internet-based rehabilitation, which combines internet technologies with physical medicine and rehabilitation, could vastly promote accessibility to professional physicians or therapists for patients, even for those residing in remote areas [18,19]. The feasibility of internet-based rehabilitation and its effect have been investigated in patients with stroke [20,21], chronic obstructive pulmonary disease [22-25], Parkinson disease [26-28], multiple sclerosis [29,30] and following knee arthroplasty [31,32].

Despite the increasing popularity of these internet-based rehabilitation programs, there is insufficient evidence to demonstrate their effectiveness for patients with knee OA. Positive results have been shown in some studies in which specific programs of internet-based rehabilitation could improve pain and physical function for patients with knee OA as compared with conventional rehabilitation [33,34]. However, different views have also been put forth, indicating no significant change in OA-related pain and physical function during long-term follow up of 12 months [35,36]. To our knowledge, there has been no meta-analysis of randomized controlled trials (RCTs) assessing the effects of internet-based rehabilitation programs on improvement of pain and physical function in patients with knee OA. Therefore, the aim of this systematic review and meta-analysis was to assess the effect of internet-based rehabilitation programs on the pain and physical function of patients with knee OA, and to evaluate the specific components (eg, exercise guidance, knee OA education) designed for each of the internet-based rehabilitation programs reported to date.

Methods

Study Protocol and Registration

All analyses were based on data from previously published studies. Thus, no ethical approval or patient consent was required. The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [37]. The a priori protocol for the review is published in the International Prospective Register of Systematic Reviews (PROSPERO): CRD42019137907.

Information Sources

The following electronic databases were searched to identify relevant studies from January 2000 to April 2020: Web of Science, MEDLINE, EMBASE, CENTRAL, Scopus, Physiotherapy Evidence Database (PEDro), CNKI, SinoMed, and WANFANG. Relevant journals, conference proceedings, and reference lists were manually searched to identify additional studies.

Search and Eligibility Criteria

Overall Search Strategy

The search was performed using a combination of the following keywords on May 1, 2020: (osteoarthritis or osteoarthrosis or cartilage or degenerative arthritis) AND (telemedicine or e-health or telehealth or telerehabilitation or internet or web or online or app or wearable or sensor) AND knee. The search strategies for each database are presented in Multimedia Appendix 1. In addition, the literature was searched manually from the reference lists of the articles identified from the search of the electronic databases. The inclusion and exclusion criteria of the studies were based on the PICO (Population, Intervention, Comparison, Outcome) method [38,39].

Studies

RCTs regarding the effect of internet-based rehabilitation programs for patients with knee OA were included in the review. The included studies were published in English or Chinese. Articles were excluded if the study was a non-RCT or nonclinical trial. Abstracts from meeting proceedings with no corresponding full article published in a peer-reviewed journal or no specific data provided even after contacting the author were excluded.

Participants

The studies involved participants aged above 18 years, who were diagnosed with knee OA by a physician or self-reported a physician diagnosis along with matching items based on the American College of Rheumatology clinical criteria [40,41], and had not undergone knee arthroplasty.

Interventions

Studies that were included in the review compared the effects of internet-based rehabilitation programs with conventional rehabilitation (eg, rehabilitation performed in the clinic or hospital) or waiting without any therapy. Internet-based rehabilitation could be the only intervention or could be combined with another form of physiotherapy. The internet-based rehabilitation programs were performed through videos or graphic knowledge demonstrations, real-time communication with physicians or therapists, and group discussions to promote the self-rehabilitation for individuals with knee OA. Rehabilitation methods include exercise, patient education, and self-management. Interventions used for participants had to be internet-based such as by email, websites, or software systems. Studies using noninternet technology support or not explicitly stating that internet technology was used to support the intervention were excluded, such as telephone, DVD, and cable television.

Outcome Measures

The main outcome measures were focused on pain (eg, the Western Ontario and McMaster [WOMAC] pain subscale, visual analog scale [VAS], Numerical Pain Rating Scale [NPRS]) and physical function (eg, WOMAC functional subscale, 30-second chair stand test, Timed Up and Go Test [TUG], and Knee Injury and OA Outcome Score [KOOS] functional subscale) for patients with knee OA. The primary outcome scale or the most representative scale was selected for analysis if multiple scales were used to evaluate the same outcome index in a study.

Search Methods for Identification of Studies

Two authors (LW and LQW) independently reviewed the search results and screened the titles, abstracts, and full texts of identified references to select potentially eligible studies, which were imported into EndNote X8 (Clarivate Analytics, Philadelphia, PA, USA).

Data Extraction and Management

Two authors (LW and LQW) completed data extraction independently and assessed the risk of bias for the included studies. A final decision was made after discussion with authors QW and CH in cases of any disagreement related to the data extraction process. Finally, the data were summarized in a previously standardized worksheet of Excel for Windows 2010. When the reported data were insufficient, we contacted the authors for more information.

The extracted data included: basic information of the study (eg, first author, year of publication, country, email address of the corresponding author); risk of bias (based on the PEDro scale) [42]; participants (overall sample size and sample size for each condition, overall mean age and the mean age for each condition, and the number of men and women); type of intervention for the experimental group (name of the program, components of the program, intervention time, delivery location); type of intervention for the control group (same as above); and outcomes (eg, the WOMAC pain and functional subscale, VAS, TUG). Outcomes reported as continuous variables are presented as the mean (SD).

Quality Assessment

Quality assessments were performed with the PEDro scale [42]. The PEDro tool is based on the Delphi List criteria, which was used to evaluate the methodological quality in this study, and is considered to be valid and reliable [42-44]. All included trial reports were checked in the PEDro database to confirm their PEDro scale score. Considering that criterion 1 was not utilized to calculate the score, the sum of the other criteria could have a maximum of 10 points. Trials with a score ≥6 points were classified as “good,” whereas those with a score ≤5 points were graded as “poor” [45]. The poor-quality studies were excluded from the analysis. The quality of studies was assessed by two authors (SX and KS) using the PEDro scale and associated notes on administration of the PEDro scale [46] independently if a score was not available in the PEDro database [47]. Any dispute was settled through discussion or with consultation of a third reviewer (QW).

Statistical Analysis

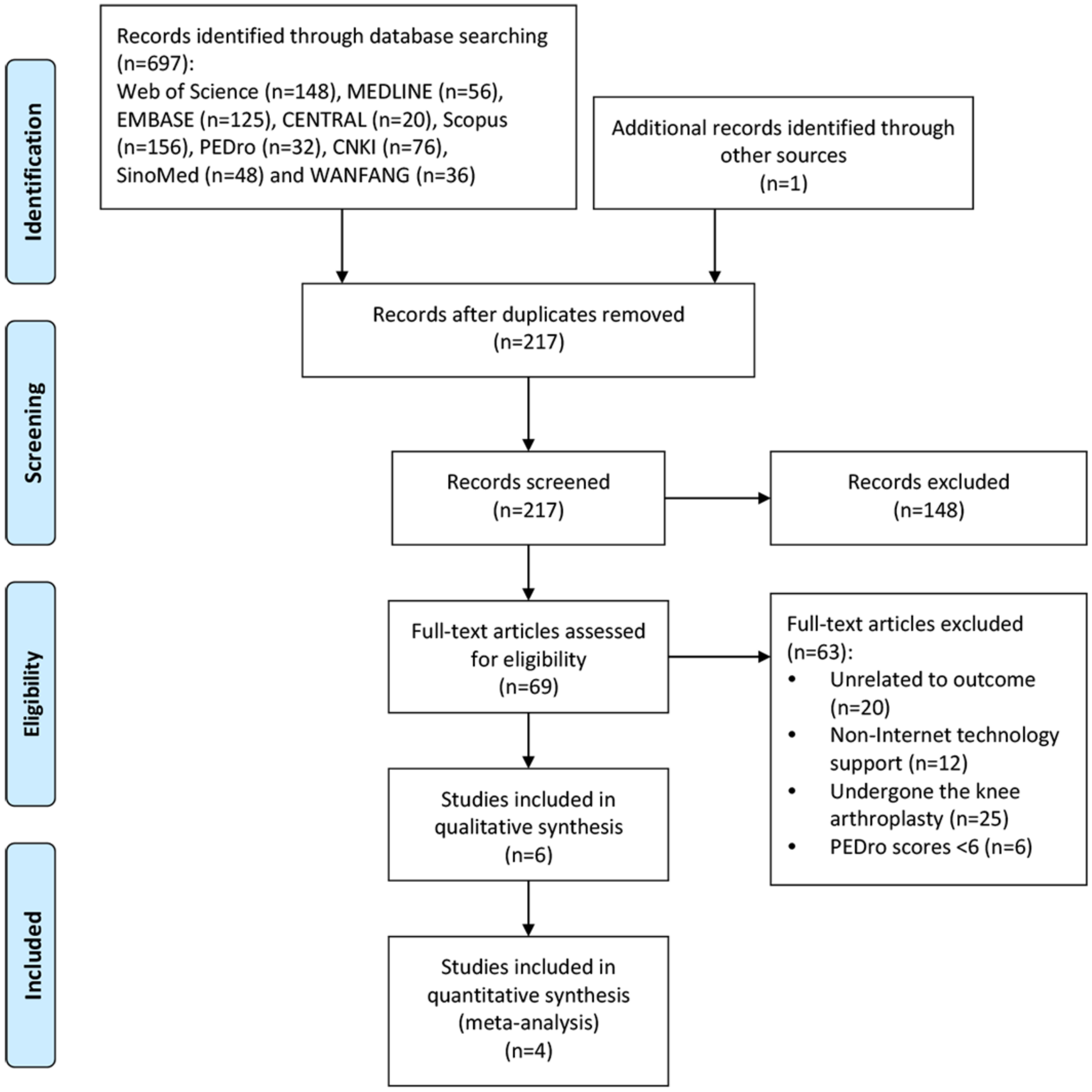
The mean (SD) of continuous outcome variables after therapy was used to calculate the total effect size via the mean difference and 95% CI. The standardized mean difference (SMD) was calculated when studies used different methods or scales to measure the same outcome. We assessed heterogeneity visually and based on the I^2 statistic [48]. The forest plots for the meta-analysis are presented along with a description of the results. A random-effects model was applied when substantial heterogeneity was observed ($P<.05$ or $I^2>50\%$); otherwise, a fixed-effects model was used [49]. Review Manager version 5.3 (Cochrane Collaboration, Copenhagen, Denmark) was employed for the statistical analyses and to produce forest plots.

Results

Search and Selection

A total of 697 publications were retrieved through electronic searching from the databases. After exclusion of the duplicated studies and irrelevant subjects via the initial screening of titles and abstracts, 12 articles were systematically reviewed with 6 studies further excluded due to low quality based on a PEDro score ≤5 points. The list of eligible studies was sent to experts in the field to confirm that no other studies could be identified. In addition, the final articles included in the systematic review and meta-analysis were determined according to the guidelines of Cochrane Handbook for Systematic Reviews of Interventions [50]. Finally, 6 studies were identified for the systematic review, 4 of which were included in the meta-analysis, involving a total of 791 patients with knee OA (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram. CENTRAL: Cochrane Central Register of Controlled Trials; EMBASE: Excerpta Medica Database; MEDLINE: Medical Literature Analysis and Retrieval System Online; PEDro: Physiotherapy Evidence Database; CNKI: China National Knowledge Infrastructure.



Study Characteristics

The baseline descriptive characteristics (country, sample size, age, and gender) of the 6 studies included in the systematic review are summarized in Table 1. Two studies were from the United States [35,51], one from Australia [33], one from China [34], one from Brazil [52], and one from the Netherlands [53]. The mean age of patients with knee OA ranged from 53.1 (SD 8.5) to 72.25 (SD 8.84) years, and all studies included both men and women.

Table 1. Baseline descriptive characteristics of studies included in the systematic review.

Reference, year, country	Patient characteristics		Comparison	Intervention	Intervention time (weeks)	Outcome
	N	Age (years), mean (SD)				
Aily et al [52][a], 2020, Brazil	20 (10 women, 10 men)	Group 1 (n=10): 54.8 (8.3); Group 2 (n=10): 53.1 (8.5)	Group 1: supervised periodized circuit training with load progression, 3 times a week	Group 2: same exercise protocol as group 1, but orientations to the exercises provided through videos on a website, DVD, or YouTube	14	VAS[b], WOMAC[c], 30-s chair stand test, 40-m fast-paced walk test, stair climb test
Huang et al [34][a], 2019, China	40 (30 women, 10 men)	Group 1 (n=20): 72.25 (8.84); Group 2 (n=20): 67.25 (10.97)	Group 1: conventional rehabilitation in the clinic	Group 2: conventional rehabilitation plus a brief GOH[d]-based intervention (educational lectures, medical suggestions, and psychotherapy)	24	WOMAC
O'Moore [33][a], 2018, Australia	69 (55 women, 14 men)	Group 1 (n=25): 59.68 (6.01); Group 2 (n=44): 63.16 (7.38)	Group 1: treatment as usual.	Group 2: iCBT[e] program for depression added to treatment as usual	10	ASES[f], WOMAC
Allen et al [35][a], 2018, United States	350 (251 women, 99 men)	Group 1 (n=140): 65.7 (10.3); Group 2 (n=68): 64.3 (12.2); Group 3 (n=142): 65.3 (11.5)	Group 1: physiotherapy (evidence-based approach); Group 2: wait without any therapy	Group 3: internet-based exercise training	48	WOMAC, 30-s chair stand, TUG[g], 2-min step test, unilateral stand time
Rini et al [51], 2015, United States	113 (91 women, 22 men)	Group 1 (n=55): 66.67 (11.02); Group 2 (n=58): 68.52 (7.65)	Group 1: wait without any therapy	Group 2: PainCOACH program through the internet	8-10	AIMS2[h] 5-item arthritis pain subscale, ASES, AIMS2 subscales relevant to lower extremity functioning, 20-item PASS[i], AIMS2 comorbidities subscale
Bossen et al [53], 2013, Netherlands	199 (129 women, 70 men)	Group 1 (n=99): 63.0 (5.4); Group 2 (n=100): 61.0 (5.9)	Group 1: waitlist control without any therapy	Group 2: Join2move, a fully automated web-based intervention without human support	48	Self-reported PA[j] (PASE[k] and ActiGraph GT3X triaxial accelerometers), KOOS[l] (functional subscale), self-perceived effect, NRS[m], ASES

[a] Included in the meta-analysis.

[b] WOMAC: Western Ontario and McMaster Universities Osteoarthritis index.

[c] VAS: visual analog scale.

[d] GOH: Guangdong Online Hospital.

[e] iCBT: internet cognitive behavior therapy.

[f] ASES: Arthritis Self Efficacy Scale.

[g] TUG: Timed Up and Go Test.

[h] AIMS2: Arthritis Impact Measurement Scale 2.

[i] PASS: Pain Anxiety Symptoms Scale.

[j] PA: physical activity.

[k] PASE: Physical Activity Scale for the Elderly.

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

[m] NRS: numeric rating scale.

Intervention Programs

The internet-based rehabilitation programs used in the included RCTs are summarized in Table 1. Various internet-based rehabilitation programs have been developed in the included studies. To promote physical fitness, Allen et al [35] from the United States developed an internet-based exercise training program (IBET) containing tailored exercises, exercise progression, video demonstrations, automated reminders, and guidance on progression for patients with OA. Participants were encouraged to complete strengthening and stretching exercises at least 3 times per week and to engage in aerobic exercises daily [35]. Similarly, Rini et al [51] from the United States developed the PainCOACH program, which is a web-based platform that offers physical, psychological, and occupational therapies. PainCOACH includes 8 modules related to cognitive or behavioral pain coping skills in a self-directed manner (eg, without therapist contact) at a frequency of one per week. Each module took 35 to 45 minutes to complete [51]. Huang and colleagues [34] from China developed an internet-based rehabilitation program for patients with knee OA comprised of three broad segments: encouragement, educational lectures, and medical issues, each of which could be completed within 20-30 min in an independent manner. In Australia, O'Moore and colleagues [33] studied the effectiveness of an internet-based cognitive-behavioral therapy (iCBT) program for older adults with knee OA. The iCBT Sadness Program consists of six online lessons assigned as regular homework and provides access to supplementary resources. Aily and colleagues [52] from Brazil allowed patients with knee OA to utilize a website or YouTube videos for rehabilitation at home 3 times a week. They also provided periodic telephone calls to motivate, clarify, and monitor the performance of patients. In addition, a behavior-graded activity program named Join2move was developed by Bossen et al [53] in the Netherlands with the aim of promoting the self-management of behaviors of patients with knee OA when they are at home or in the community setting. The intervention period ranged from 8-10 weeks [51] to 48 weeks [35,53] (Table 1).

Risk of Bias

All 6 studies included in the review scored greater than 6 points on the PEDro scale (Table 2). Even though the greatest risk of bias lies in the nonblinding of participants, in general, the total score of the 6 studies showed high methodological rigor [54,55] despite the fact that all studies had inadequate blinding of participants and therapists, and 2 studies had inadequate blinding of outcome assessors. One study did not provide adequate follow up. Two studies did not perform an intention-to-treat analysis. Overall, the methodological quality of the included studies was assessed as "good."

Table 2. Assessment of methodological quality using the PEDro scale.

Quality metric	Aily et al [52]	Huang et al [34]	O'Moore et al [33]	Allen et al [35]	Rini et al [51]	Bossen et al [53]
Eligibility criteria	Yes	Yes	Yes	Yes	Yes	Yes
Random allocation	Yes	Yes	Yes	Yes	Yes	Yes
Concealed allocation	Yes	Yes	Yes	Yes	Yes	Yes
Baseline comparability	Yes	Yes	Yes	Yes	Yes	Yes
Blinded subjects	No	No	No	No	No	No
Blinded therapists	No	No	No	No	No	No
Blinded assessors	No	Yes	Yes	Yes	Yes	No
Adequate follow up	Yes	Yes	Yes	Yes	Yes	No
Intention-to-treat analysis	No	No	Yes	Yes	Yes	Yes
Between-group comparisons	Yes	Yes	Yes	Yes	Yes	Yes
Point estimates and variability	Yes	Yes	Yes	Yes	Yes	Yes
Total score[a]	6	7	8	8	8	6
Quality assessment	Good	Good	Good	Good	Good	Good

[a] Eligibility criteria did not contribute to the total score: 1=yes, 0=no.

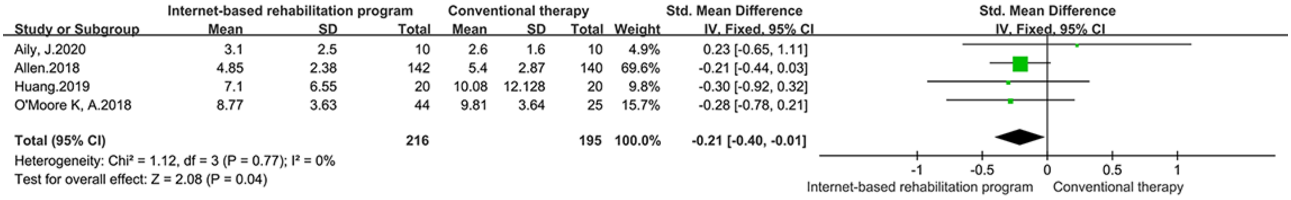
Outcomes of Interest

Pain

Pain is the most disabling symptom for patients with knee OA [5]. Six studies included in the review evaluated the effect of internet-based rehabilitation on osteoarthritic pain. The WOMAC pain subscale was used in 4 studies [33-35,52], with scores on the pain subscale ranging from 0 (no dysfunction) to 20 (maximum dysfunction) based on a 5-point Likert format, or ranging from 0 (no dysfunction) to 50 (maximum dysfunction) with the 11-box numerical rating scale format [56]. In the study of Aily et al [52], the WOMAC and VAS were both used to assess pain. The VAS is a 100-mm line and the participants are required to place a mark between the left side (0, representing "no pain") and the right side (100, representing "the worst pain imaginable"). The 10-NPRS and the Arthritis Impact Measurement Scale 2 (AIMS2) were used in the other two studies, respectively [51,53]. The NPRS is scored in a similar manner to the VAS, except that the NPRS is scored from 0 to 10 (0 means no pain and 10 means the worst possible pain). AIMS2 is comprised of a 5-item arthritis pain subscale indicating the severity of arthritis pain (1 means severe and 5 means none) and the frequency of severe pain [57]. A study that did not apply the WOMAC function subscale measures showed a significant improvement in patients with knee OA after 3 months of internet-based rehabilitation based on the NPRS, but no significant change was observed after 12 months compared with the control group [53]. In another study using the AIMS2 pain subscale, the pain after 8-10 weeks of internet-based rehabilitation was found to be significantly improved in women but not in men compared with that of the control group [51].

The meta-analysis on the effect of internet-based rehabilitation on osteoarthritic pain as measured by the WOMAC pain subscale contained 4 independent studies, involving a total of 411 participants [33-35,52]. We did not find evidence of significant heterogeneity among these studies (I^2=0%, *P*=.77); therefore, a fixed-effects model was used. The meta-analysis showed that internet-based rehabilitation could significantly reduce the pain of patients with knee OA compared with conventional rehabilitation as assessed by the WOMAC pain subscale (SMD –0.21, 95% CI −0.4 to –0.01, *P*=.04; Figure 2).

Figure 2. Forest plot of included studies comparing the effect of the internet-based intervention and conventional rehabilitation on pain according to the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) pain subscale.

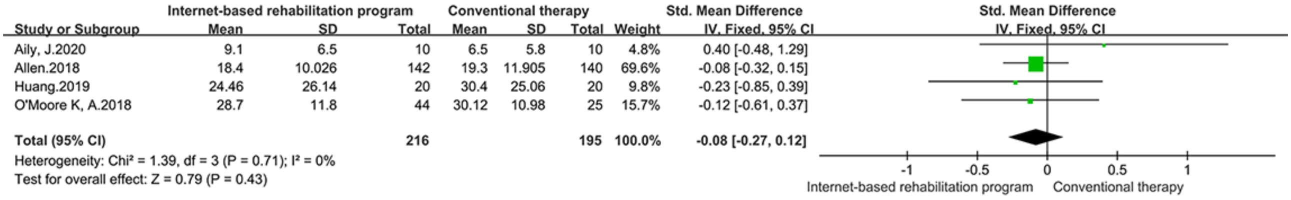


Physical Function

Improving patients' functional conditions is the objective of rehabilitation for patients with knee OA. Physical function was assessed in 4 studies using the WOMAC function subscale [33-35,52]. The total score of the WOMAC function subscale (17 items) ranges from 0 (no dysfunction) to 68 (maximum dysfunction) with a 5-point Likert response format or from 0 (no dysfunction) to 170 (maximum dysfunction) with the 11-box numerical rating scale format [56]. One study assessed physical function using the KOOS function subscale [53]. The KOOS is a self-administered questionnaire to assess functional status regarding the patient's knee problems on a 5-point Likert scale [58]. In another study, the AIMS2 subscales were used to evaluate the function of the lower extremities [51]. The TUG, 40-meter fast-paced walk test, 2-minute step test, unilateral stand time, and 30-second chair stand test were used to assess physical function for patients with knee OA in two studies [35,52]. The studies that did not apply the WOMAC function subscale measures showed that internet-based rehabilitation could not significantly improve physical function compared with the control group [35,52], even compared with the waitlist group as a control [35]. Only one study suggested that the physical function could be significantly improved after 3 months of internet-based rehabilitation compared with the waitlist group. However, the beneficial effect did not last after 12 months [53].

The 4 studies included in the meta-analysis on the effect of internet-based rehabilitation on function as measured by the WOMAC function subscale involved a total of 411 participants [33-35,52]. The fixed-effects model was used in this analysis owing to the low heterogeneity (I^2=0%, *P*=.71). The results indicated that internet-based rehabilitation could not significantly improve the physical function of patients with knee OA compared with the control group according to the WOMAC function subscale (SMD –0.08, 95% CI −0.27 to 0.12, *P*=.43; Figure 3).

Figure 3. Forest plot of included studies comparing the effect of the internet-based intervention and conventional rehabilitation on physical function based on the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) function subscale.



Discussion

Principal Findings

This systematic review and meta-analysis investigated whether internet-based rehabilitation programs could effectively improve the pain and physical function in patients with knee OA. The findings showed that internet-based rehabilitation could significantly improve the pain of patients with knee OA but not the physical function. Qualitative synthesis was performed for 6 studies and the meta-analysis was performed for 4 studies, comprising a total of 791 patients with knee OA. These trials showed good methodological quality as assessed by high PEDro scores (>6). However, only a very small number of studies could be included in the review.

Knee OA often causes pain, which is a significant reason for patients to be admitted to the hospital. This meta-analysis showed that internet-based rehabilitation could assist patients with knee OA to self-manage and even relieve their pain after they are discharged from the hospital. The programs such as IBET and the web-based intervention Join2move were demonstrated to be effective for pain reduction in patients with knee OA [35,53]. The IBET program focused on exercise interventions tailored to patients' needs, and played a role in pain control [35]. Similarly, Join2move, developed by Bossen and colleagues [53], adopts a behavior-graded activity program to assist patients with OA to gradually increase their daily activities in a fixed amount of time [53]. These internet-based rehabilitation programs could combine various interventions based on the patient's behavioral, psychological, family, and social factors, which can be carried out at home or in the community setting to alleviate osteoarthritic pain.

Furthermore, knee OA could lead to the decline of patients' physical function such as walking, shopping, and housework [59]. This meta-analysis showed that internet-based rehabilitation could not significantly improve the physical function of patients with knee OA compared with conventional rehabilitation. Allen et al [35] reported that patients who underwent 12 months of internet-based strengthening and stretching exercises at least 3 times per week did not obtain a significant improvement of physical function compared with those who received conventional rehabilitation through face-to-face supervised exercise. Even though the patients' physical function could be improved after 3 months of the Join2move internet-based rehabilitation, the positive effects were not detectable at follow up of 12 months compared with the waitlist group [53]. It was postulated that the undetectable improvement after internet-based rehabilitation might be due to the fact that the recruited participants often had better baseline physical function and could accomplish the tasks assigned by the programs [60,61]. Alternatively, we also speculate that functional improvement may require a longer-term intervention and more intervention forms that can integrate daily life factors and improve lifestyle function, rather than simply the external exercise components. These results are comparable with the results of other meta-analyses [62,63]. Wang et al [62] showed that a telerehabilitation program (eg, telephone counseling/coaching, video conferencing) could be effective for pain control but not for functional improvement in patients after knee OA replacement surgery. Another systematic review also showed that the improvements in physical function were not significant for patients with knee OA through telerehabilitation exercise compared with either control or waitlist groups [63].

It is critical for the designers or health care providers to develop appropriate modules comprised in the internet-based rehabilitation programs. In this systematic review, we found that the programs for knee OA possessed some common modules such as exercise guidance, psychological intervention, knee OA education, and cognitive behavior management. In addition, the different programs were manifested through the specific modules. For example, the PainCOACH program focused on the behavioral and cognitive management of pain control for patients with knee OA, whereas the Join2move program aimed to enhance the physical function [51,53]. In the future, comprehensive and personalized modules will need to be developed to achieve the integration of facilities and patients in the community or home setting, and to monitor the safety and progress via wearable devices when performing exercises or behavior management for patients with knee OA. The modules designed for each individual can be personalized based on big data analysis collected from the wearable devices.

Limitations

There are several limitations to this study. First, there were only 6 studies that could be included in the systematic review and only 4 studies that were eligible to be included in the meta-analysis. Thus, more high-quality RCTs with larger sample sizes in this field are needed. Second, the included participants were mainly patients with knee OA who had not undergone arthroplasty or other surgical interventions. Third, the outcome measures used to assess pain and physical function included in the studies of internet-based rehabilitation were subjective. Fourth, analyses of moderator variables on the effects of the internet-based rehabilitation programs (eg, age, gender, sample size) were not performed. Fifth, considering the diversity of outcome indicators and the small number of included studies, only the studies using the WOMAC scale were included in the meta-analysis to ensure the reliability of the study and comparison. We plan to update this systematic review and meta-analysis with the increase of research on internet-based rehabilitation for knee OA in the future. Finally, only 2 studies included a follow-up period of 12 months, indicating a lack of assessments on the long-term effects of internet-based rehabilitation.

Conclusion

Internet-based rehabilitation is a promising strategy for patients with knee OA to obtain access to rehabilitation guidance and monitoring at home or in the community setting. The results of this systematic review and meta-analysis indicate that internet-based rehabilitation programs involving personalized modules could improve the pain but not the physical function of patients with knee OA compared with conventional rehabilitation. More high-quality studies with large samples are needed, with a focus on the long-term outcomes of internet-based rehabilitation for patients with knee OA.

Acknowledgments

The authors would like to acknowledge the contributions of two postdoctoral fellows at the Department of Rehabilitation Medicine and Physiotherapy from Sichuan University, Siyi Zhu and Yi Li, for assisting with the literature search and acquisition of articles, and for providing advice on statistical analysis for this systematic review. We extend special thanks to Diya Yang for her encouragement and assistance in the writing of the manuscript. This study was supported by grants from the National Natural Science Foundation of China (81572236) and the Post-Doctoral Research Project (2018HXBH080) of West China Hospital of Sichuan University. These funding sources were not involved in the literature review, systematic review, meta-analysis, and writing of the report.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 17 KB - jmir_v23i1e21542_app1.docx]

References

1. GBD 2017 DALYsHALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1859-1922 [FREE Full text] [doi: 10.1016/S0140-6736(18)32335-3] [Medline: 30415748]
2. Briggs AM, Cross MJ, Hoy DG, Sànchez-Riera L, Blyth FM, Woolf AD, et al. Musculoskeletal Health Conditions Represent a Global Threat to Healthy Aging: A Report for the 2015 World Health Organization World Report on Ageing and Health. Gerontologist 2016 Apr;56(Suppl 2):S243-S255. [doi: 10.1093/geront/gnw002] [Medline: 26994264]
3. Bowden JL, Hunter DJ, Deveza LA, Duong V, Dziedzic KS, Allen KD, et al. Core and adjunctive interventions for osteoarthritis: efficacy and models for implementation. Nat Rev Rheumatol 2020 Aug;16(8):434-447. [doi: 10.1038/s41584-020-0447-8] [Medline: 32661322]
4. Prieto-Alhambra D, Judge A, Javaid MK, Cooper C, Diez-Perez A, Arden NK. Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. Ann Rheum Dis 2014 Sep;73(9):1659-1664 [FREE Full text] [doi: 10.1136/annrheumdis-2013-203355] [Medline: 23744977]
5. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019 Apr 27;393(10182):1745-1759. [doi: 10.1016/S0140-6736(19)30417-9] [Medline: 31034380]
6. Murphy LB, Sacks JJ, Brady TJ, Hootman JM, Chapman DP. Anxiety and depression among US adults with arthritis: prevalence and correlates. Arthritis Care Res (Hoboken) 2012 Jul;64(7):968-976. [doi: 10.1002/acr.21685] [Medline: 22550055]
7. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
8. de Rooij M, van der Leeden M, Cheung J, van der Esch M, Häkkinen A, Haverkamp D, et al. Efficacy of Tailored Exercise Therapy on Physical Functioning in Patients With Knee Osteoarthritis and Comorbidity: A Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2017 Jun;69(6):807-816. [doi: 10.1002/acr.23013] [Medline: 27563831]
9. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008 Feb;16(2):137-162 [FREE Full text] [doi: 10.1016/j.joca.2007.12.013] [Medline: 18279766]
10. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SMA, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019 Nov;27(11):1578-1589 [FREE Full text] [doi: 10.1016/j.joca.2019.06.011] [Medline: 31278997]
11. Rosen MJ. Telerehabilitation. Telemed J E Health 2004;10(2):115-117. [doi: 10.1089/tmj.2004.10.115] [Medline: 15319039]
12. Committee on Evaluating Clinical Applications of Telemedicine. In: Field MJ, editor. Telemedicine: A Guide to Assessing Telecommunications in Health Care. Washington DC: National Academies Press; 1996:136-137.
13. Odole AC, Ojo OD. A Telephone-based Physiotherapy Intervention for Patients with Osteoarthritis of the Knee. Int J Telerehabil 2013;5(2):11-20 [FREE Full text] [doi: 10.5195/ijt.2013.6125] [Medline: 25945214]
14. Burns RB, Crislip D, Daviou P, Temkin A, Vesmarovich S, Anshutz J, et al. Using telerehabilitation to support assistive technology. Assist Technol 1998;10(2):126-133. [doi: 10.1080/10400435.1998.10131970] [Medline: 10339280]
15. Lee AC, Davenport TE, Randall K. Telehealth Physical Therapy in Musculoskeletal Practice. J Orthop Sports Phys Ther 2018 Oct;48(10):736-739. [doi: 10.2519/jospt.2018.0613] [Medline: 30270782]
16. Russell TG. Physical rehabilitation using telemedicine. J Telemed Telecare 2007;13(5):217-220. [doi: 10.1258/135763307781458886] [Medline: 17697506]
17. Galea MD. Telemedicine in Rehabilitation. Phys Med Rehabil Clin N Am 2019 May;30(2):473-483. [doi: 10.1016/j.pmr.2018.12.002] [Medline: 30954160]
18. Burridge JH, Lee ACW, Turk R, Stokes M, Whitall J, Vaidyanathan R, et al. Telehealth, Wearable Sensors, and the Internet: Will They Improve Stroke Outcomes Through Increased Intensity of Therapy, Motivation, and Adherence to Rehabilitation Programs? J Neurol Phys Ther 2017 Jul;41(Suppl 3):S32-S38. [doi: 10.1097/NPT.0000000000000183] [Medline: 28628594]
19. Hallensleben C, van Luenen S, Rolink E, Ossebaard HC, Chavannes NH. eHealth for people with COPD in the Netherlands: a scoping review. Int J Chron Obstruct Pulmon Dis 2019;14:1681-1690. [doi: 10.2147/COPD.S207187] [Medline: 31440044]
20. Tchero H, Tabue Teguo M, Lannuzel A, Rusch E. Telerehabilitation for Stroke Survivors: Systematic Review and Meta-Analysis. J Med Internet Res 2018 Oct 26;20(10):e10867 [FREE Full text] [doi: 10.2196/10867] [Medline: 30368437]
21. Sarfo FS, Adamu S, Awuah D, Sarfo-Kantanka O, Ovbiagele B. Potential role of tele-rehabilitation to address barriers to implementation of physical therapy among West African stroke survivors: A cross-sectional survey. J Neurol Sci 2017 Oct 15;381:203-208 [FREE Full text] [doi: 10.1016/j.jns.2017.08.3265] [Medline: 28991682]

22. Zanaboni P, Dinesen B, Hjalmarsen A, Hoaas H, Holland AE, Oliveira CC, et al. Long-term integrated telerehabilitation of COPD Patients: a multicentre randomised controlled trial (iTrain). BMC Pulm Med 2016 Aug 22;16(1):126 [FREE Full text] [doi: 10.1186/s12890-016-0288-z] [Medline: 27549782]
23. Paneroni M, Colombo F, Papalia A, Colitta A, Borghi G, Saleri M, et al. Is Telerehabilitation a Safe and Viable Option for Patients with COPD? A Feasibility Study. COPD 2015 Apr;12(2):217-225. [doi: 10.3109/15412555.2014.933794] [Medline: 25093419]
24. Holland AE, Cox NS. Telerehabilitation for COPD: Could pulmonary rehabilitation deliver on its promise? Respirology 2017 May;22(4):626-627. [doi: 10.1111/resp.13028] [Medline: 28345290]
25. Tabak M, Vollenbroek-Hutten MM, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]
26. Dias AE, Limongi JCP, Hsing WT, Barbosa ER. Telerehabilitation in Parkinson's disease: Influence of cognitive status. Dement Neuropsychol 2016;10(4):327-332 [FREE Full text] [doi: 10.1590/s1980-5764-2016dn1004012] [Medline: 29213477]
27. Gandolfi M, Geroin C, Dimitrova E, Boldrini P, Waldner A, Bonadiman S, et al. Virtual Reality Telerehabilitation for Postural Instability in Parkinson's Disease: A Multicenter, Single-Blind, Randomized, Controlled Trial. Biomed Res Int 2017;2017:7962826. [doi: 10.1155/2017/7962826] [Medline: 29333454]
28. Constantinescu G, Theodoros D, Russell T, Ward E, Wilson S, Wootton R. Assessing disordered speech and voice in Parkinson's disease: a telerehabilitation application. Int J Lang Commun Disord 2010;45(6):630-644. [doi: 10.3109/13682820903470569] [Medline: 20102257]
29. Finkelstein J, Liu J. Usability of Telerehabilitation System Supporting Multipronged Exercise in Patients with Multiple Sclerosis. Stud Health Technol Inform 2018;251:281-284. [Medline: 29968658]
30. Khan F, Amatya B, Kesselring J, Galea MP. Telerehabilitation for persons with multiple sclerosis. A Cochrane review. Eur J Phys Rehabil Med 2015 Jun;51(3):311-325 [FREE Full text] [Medline: 25943744]
31. Tousignant M, Boissy P, Corriveau H, Moffet H, Cabana F. In-home telerehabilitation for post-knee arthroplasty: a pilot study. Int J Telerehabil 2009;1(1):9-16 [FREE Full text] [doi: 10.5195/ijt.2009.5997] [Medline: 25945158]
32. Tousignant M, Moffet H, Boissy P, Corriveau H, Cabana F, Marquis F. A randomized controlled trial of home telerehabilitation for post-knee arthroplasty. J Telemed Telecare 2011;17(4):195-198. [doi: 10.1258/jtt.2010.100602] [Medline: 21398389]
33. O'Moore KA, Newby JM, Andrews G, Hunter DJ, Bennell K, Smith J, et al. Internet Cognitive-Behavioral Therapy for Depression in Older Adults With Knee Osteoarthritis: A Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2018 Jan;70(1):61-70. [doi: 10.1002/acr.23257] [Medline: 28426917]
34. Huang Z, Zhong X, Xie Z, Li T. The feasibility and effectiveness of telemedicine for knee osteoarthritis in disease management: A randomised control trial. In: Ann Rheum Dis (Supplement 2). 2019 Presented at: Annual European Congress of Rheumatology (EULAR); June 12-15, 2019; Madrid, Spain. [doi: 10.1136/annrheumdis-2019-eular.6474]
35. Allen K, Arbeeva L, Callahan L, Golightly Y, Goode A, Heiderscheit B, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis Cartilage 2018 Mar;26(3):383-396 [FREE Full text] [doi: 10.1016/j.joca.2017.12.008] [Medline: 29307722]
36. Kloek CJJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a Blended Physical Therapist Intervention in People With Hip Osteoarthritis, Knee Osteoarthritis, or Both: A Cluster-Randomized Controlled Trial. Phys Ther 2018 Jul 01;98(7):560-570 [FREE Full text] [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
37. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
38. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012 Feb 09;1:2 [FREE Full text] [doi: 10.1186/2046-4053-1-2] [Medline: 22587842]
39. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011 Apr;64(4):395-400. [doi: 10.1016/j.jclinepi.2010.09.012] [Medline: 21194891]
40. Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991 May;34(5):505-514. [doi: 10.1002/art.1780340502] [Medline: 2025304]
41. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986 Aug;29(8):1039-1049. [doi: 10.1002/art.1780290816] [Medline: 3741515]
42. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003 Aug;83(8):713-721. [Medline: 12882612]

43. de Morton NA. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. Aust J Physiother 2009;55(2):129-133. [doi: 10.1016/s0004-9514(09)70043-1] [Medline: 19463084]
44. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998 Dec;51(12):1235-1241. [doi: 10.1016/s0895-4356(98)00131-0] [Medline: 10086815]
45. Foley NC, Bhogal SK, Teasell RW, Bureau Y, Speechley MR. Estimates of quality and reliability with the physiotherapy evidence-based database scale to assess the methodology of randomized controlled trials of pharmacological and nonpharmacological interventions. Phys Ther 2006 Jun;86(6):817-824. [Medline: 16737407]
46. PEDro scale. PEDro: Physiotherapy Evidence Database. URL: https://pedro.org.au/english/resources/pedro-scale/ [accessed 2020-03-01]
47. PEDro, The Physiotherapy Evidence Database. URL: https://pedro.org.au/ [accessed 2020-03-01]
48. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019 Oct 03;10:ED000142. [doi: 10.1002/14651858.ED000142] [Medline: 31643080]
49. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
50. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Training. URL: https://training.cochrane.org/handbook/current [accessed 2020-03-03]
51. Rini C, Porter LS, Somers TJ, McKee DC, DeVellis RF, Smith M, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]
52. Aily J, Castilho de Almeida A, da Silva Ribeiro G, de Noronha M, Mattiello S. Is a periodized circuit training delivered by telerehabilitation effective for patients with knee osteoarthritis? a phase i randomized controlled trial. In: Osteoarthritis and Cartilage. 2020 Apr Presented at: OARSI World Congress on Osteoarthritis - Promoting Clinical and Basic Research in Osteoarthritis; April 30-May 3, 2020; Vienna, Austria p. S468-S469. [doi: 10.1016/j.joca.2020.02.734]
53. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
54. Fitzgerald GK, Hinman RS, Zeni J, Risberg MA, Snyder-Mackler L, Bennell KL. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of rehabilitation interventions for osteoarthritis. Osteoarthritis Cartilage 2015 May;23(5):803-814 [FREE Full text] [doi: 10.1016/j.joca.2015.03.013] [Medline: 25952351]
55. Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in Physical Therapy Trials and Its Association with Treatment Effects: A Meta-epidemiological Study. Am J Phys Med Rehabil 2017 Jan;96(1):34-44. [doi: 10.1097/PHM.0000000000000521] [Medline: 27149591]
56. Ackerman I. Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Aust J Physiother 2009;55(3):213. [doi: 10.1016/s0004-9514(09)70088-1] [Medline: 19736686]
57. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992 Jan;35(1):1-10. [doi: 10.1002/art.1780350102] [Medline: 1731806]
58. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
59. Collins NJ, Hart HF, Mills KAG. Osteoarthritis year in review 2018: rehabilitation and outcomes. Osteoarthritis Cartilage 2019 Mar;27(3):378-391 [FREE Full text] [doi: 10.1016/j.joca.2018.11.010] [Medline: 30529739]
60. Allen KD, Arbeeva L, Callahan LF, Golightly YM, Goode AP, Heiderscheit BC, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis Cartilage 2018 Mar;26(3):383-396 [FREE Full text] [doi: 10.1016/j.joca.2017.12.008] [Medline: 29307722]
61. Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S, et al. Effects of internet-based pain coping skills training before home exercise for individuals with hip osteoarthritis (HOPE trial): a randomised controlled trial. Pain 2018 Sep;159(9):1833-1842. [doi: 10.1097/j.pain.0000000000001281] [Medline: 29794609]
62. Wang X, Hunter DJ, Vesentini G, Pozzobon D, Ferreira ML. Technology-assisted rehabilitation following total knee or hip replacement for people with osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2019 Nov 03;20(1):506 [FREE Full text] [doi: 10.1186/s12891-019-2900-x] [Medline: 31679511]
63. Schäfer AGM, Zalpour C, von Piekartz H, Hall TM, Paelke V. The Efficacy of Electronic Health-Supported Home Exercise Interventions for Patients With Osteoarthritis of the Knee: Systematic Review. J Med Internet Res 2018 Apr 26;20(4):e152 [FREE Full text] [doi: 10.2196/jmir.9465] [Medline: 29699963]

Abbreviations

AIMS2: Arthritis Impact Measurement Scale 2
IBET: internet-based exercise training
iCBT: internet cognitive behavior therapy
KOOS: Knee Injury and Osteoarthritis Outcome Score
NPRS: Numeric Pain Rating Scale
OA: osteoarthritis
PEDro: Physiotherapy Evidence Database
RCT: randomized controlled trial
SMD: standardized mean difference
TUG: The Timed Up and Go Test
VAS: The visual analog scale
WOMAC: The Western Ontario and McMaster Universities Osteoarthritis

---

Edited by G Eysenbach; submitted 17.06.20; peer-reviewed by J Vermeir, TW Chien; comments to author 01.07.20; revised version received 29.08.20; accepted 15.11.20; published 05.01.21.

Please cite as:
Xie SH, Wang Q, Wang LQ, Wang L, Song KP, He CQ
Effect of Internet-Based Rehabilitation Programs on Improvement of Pain and Physical Function in Patients with Knee Osteoarthritis: Systematic Review and Meta-analysis of Randomized Controlled Trials
J Med Internet Res 2021;23(1):e21542
URL: https://www.jmir.org/2021/1/e21542
doi:10.2196/21542
PMID:33399542

Review

Assessment of Psychological Distress in Adults With Type 2 Diabetes Mellitus Through Technologies: Literature Review

Giulia Bassi[1*], PhD; Silvia Gabrielli[2*], PhD; Valeria Donisi[2*], PhD; Sara Carbone[2*], MSc; Stefano Forti[2*], PhD; Silvia Salcuni[1*], PhD

[1]Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy

[2]Fondazione Bruno Kessler, Trento, Italy

[*]all authors contributed equally

Corresponding Author:
Giulia Bassi, PhD
Department of Developmental Psychology and Socialization
University of Padova
Via Venezia 12
Padova, 35131
Italy
Phone: 39 393477334405
Email: giulia.bassi@phd.unipd.it

Abstract

Background: The use of technological devices can support the self-management of individuals with type 2 diabetes mellitus (T2DM), particularly in addressing psychological distress. However, there is poor consistency in the literature regarding the use of psychological instruments for the web-based screening of patients' psychological distress and subsequent monitoring of their psychological condition during digital interventions.

Objective: This study aims to review previous literature on the types of psychological instruments delivered in digital interventions for assessing depression, anxiety, and stress in patients with T2DM.

Methods: The literature review was conducted using the PsycINFO, CINAHL and PubMed databases, in which the following terms were considered: *diabetes mellitus, measure, assessment, self-care, self-management, depression, anxiety, stress, technology, eHealth, mobile health, mobile phone, device,* and *smartphone.*

Results: In most studies, psychological assessments were administered on paper. A few studies deployed self-reporting techniques employing automated telephonic assessment, a call system for screening and monitoring patients' conditions and preferences, or through telephone interviews via interactive voice response calls, a self-management support program leveraging tailored messages and structured emails. Other studies used simple telephone interviews and included the use of apps for tablets and smartphones to assess the psychological well-being of patients. Finally, some studies deployed mood rating scales delivered through tailored text message–based support systems.

Conclusions: The deployment of appropriate psychological tools in digital interventions allows researchers and clinicians to make the screening of anxiety, stress, and depression symptoms faster and easier in patients with T2DM. Data from this literature review suggest that mobile health solutions may be preferred tools to use in such digital interventions.

(J Med Internet Res 2021;23(1):e17740) doi:10.2196/17740

KEYWORDS

type 2 diabetes mellitus; technology assessment; psychological distress; technology; review; mobile phone

Introduction

Background

After the diagnosis of type 2 diabetes mellitus (T2DM), people have to follow new lifestyles by changing their physical and psychological habits. According to several studies, T2DM is associated with significant psychological impairments, particularly depression, anxiety, and stress [1-4]. The prevalence of depression (22.4%) and anxiety (32%) in individuals with T2DM is considerably higher than that in the general population (10%), with a negative impact on the disease itself [5]. The prevalence of depressive symptoms in older adults with T2DM can be as high as 79.4% [6]. Indeed, some studies have reported that high levels of anxiety, depression, and stress tend to cause impairment in health-related quality of life and poor disease outcomes [6-9]. Factors that affect the quality of life of patients with T2DM include medical comorbidities [10], older age [11], female gender [12], and living in rural areas [10]. For instance, authors have suggested that poor glycemic control is associated with the onset of depressive symptoms [7,8] as well as diabetes distress [9], and these symptoms can improve with better glycemic control. In particular, diabetes distress is different from psychological distress, as the latter refers to a general state of emotional disturbance consisting of symptoms of depression and anxiety [13], whereas diabetes distress or diabetes-specific distress is a specific term that describes an emotional state where individuals experience stress, guilt or denial, and the burden of self-management due to diabetes itself. If such symptoms remain untreated, mild diabetes distress can result in severe diabetes distress and/or depression [14,15]. Moreover, individuals with T2DM who have reported symptoms of anxiety should be motivated to self-monitor glycemic levels while they are symptomatic [6]. Indeed, symptoms of anxiety during euglycemia—a normal level of sugar in the blood—would be suggestive of an anxiety disorder [6]. People with T2DM also presented with more stress symptoms compared with individuals who do not have this chronic disease [16]. Several studies have highlighted that stress symptoms can interact with the endocrine system, thereby involving physical attitude and nutritional behaviors, by increasing the capacity to control blood glucose [17]. Individuals with T2DM require continuous monitoring by health care professionals regarding the organic effects of diabetes; conversely, anxiety, depression, and especially stress symptoms often remain unrecognized and therefore untreated [18]. Previous studies have shown that these psychological symptoms increase the risk of more negative outcomes related to diabetes, such as glycemic control and impaired cardiovascular functioning [19]. Indeed, authors have suggested integrating psychological and medical care to address psychological symptoms and unhealthy habits (ie, sedentary lifestyle, poor diet), which often accompany depression, anxiety, and stress, as they seem to implicate benefits regarding the disease itself [6], the patients' quality of life, and their psychological well-being. Therefore, early detection and prompt treatment of anxiety, depression, and stress symptoms can lead to a better medical prognosis and a better quality of life for individuals with T2DM. For instance, patients with this disease who have received psychological interventions have shown an increase in satisfaction with treatment [20]. Therefore, as a first step, it is important to identify valid tools that can assess the levels of anxiety, depression, and stress related to T2DM management to define the appropriate treatment. Within this framework, technological devices for supporting self-care in patients with T2DM are increasing globally [21]. In particular, these instruments are mainly focused on monitoring blood glucose and physical activities through several technological devices. The use of technological apps helps support people with T2DM in the management of their psychological distress and stress. However, although studies to date have shown promising results in the use of smartphone apps for diabetes management, there are inconsistent findings regarding the type of technological devices through which the best psychological instruments, developed as paper-and-pencil tools, should be delivered to achieve a more in-depth screening of patients' psychological distress and thereby better outline the psychological intervention [22].

Objective

This paper intends to identify and outline the types of technological devices through which psychological instruments should be delivered for an accurate assessment of diabetes-related psychological symptoms, such as stress, depression, and anxiety, in which technology represents support aid for self-care and self-management of T2DM. Indeed, a precise assessment of psychological symptoms through technologies in the field of diabetes is crucial to identify and understand problematic areas to better manage the disease itself, reduce such symptoms, and facilitate the management of the disease itself. Although the number of psychological instruments designed specifically for diabetes has increased, reviews of psychological instruments integrated into technological devices for assessing psychological symptoms are generally scarce among patients with T2DM. More specifically, the aim of this review is two-fold: (1) to summarize the types of technological devices used to administer self-report questionnaires for the assessment of psychological symptoms among individuals with T2DM, with a specific focus on the efficacy and usability of these tools and (2) to summarize the principal instruments as well as their psychometric characteristics to assess psychological symptoms related to T2DM, with a focus on symptoms of anxiety, stress, and depression, through the use of technology.

Methods

Study Design

This literature review sheds light on the types of technological devices through which psychological instruments were used to assess psychological distress among adults with T2DM. This review was conducted through the academic databases PubMed (360 articles), PsycINFO (165 articles), and CINAHL (239 articles) in which the following terms and their derivatives were considered during the search: diabetes mellitus, measure, assessment, self-care, self-management, depression, anxiety, stress, technology, eHealth, mobile health, mobile phone, device, and smartphone. More specifically, the terms self-care and self-management have been used as search words to find articles in which technological devices are mentioned. Moreover, according to the Cambridge Dictionary, the term device refers to a machine, for instance, a phone or a computer, which can be used to connect to the internet [23]. In this review, these search terms were used to identify devices suitable for assessing psychological distress in T2DM. We began the review with an examination of the types of technologies through which psychological questionnaires were administered and continued by examining the principal instruments for assessing psychological symptoms related to T2DM. We then concluded the review by providing directions for future work and clinical implications.

Inclusion Criteria

The studies included in the review were in line with the following inclusion criteria: (1) presence of technological support for mental health assessment delivered to patients with T2DM; (2) studies with at least 60% of participants with T2DM, in order to have most people with only T2DM, the target population of this study; (3) studies that focused on depression, anxiety, stress, or other psychological symptoms in patients with T2DM; and (4) samples comprising adults aged between 18 and 70 years who may present with psychological distress, with a focus on depression, anxiety, and/or stress symptoms related to T2DM.

Exclusion Criteria

Studies that met any of the following criteria were excluded: (1) absence of technological support for mental health assessment; (2) studies involving patients with other chronic conditions or primary diseases (eg, cardiovascular disease, cardiomyopathy, chronic kidney disease) or psychiatric disorders, according to the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [24]; (3) presence of individuals with the risk of T2DM onset; (4) studies that focused only on monitoring glycemic control or physical activity in patients with T2DM; (5) presence of cognitive dysfunction in patients with T2DM; (6) samples based only on individuals with type 1 diabetes mellitus; (7) studies that took into account samples of children and adolescents and/or parents supporting their children with diabetes; and (8) studies that focused on pregnant women with diabetes.

Results

Included Studies

On the basis of the inclusion and exclusion criteria, 17 articles were eligible for the literature review.

Types of Technologies Through Which Tools Were Administered

Most studies (6 articles) were conducted in paper-based format—at baseline and follow-up—through written questionnaires provided by the staff or employing interviews before the technological intervention to evaluate the psychological distress and health-related quality of life of patients with T2DM [25-30]. Furthermore, 4 studies administered self-reports through the automated telephonic assessment (ATA), a call system for screening and monitoring that is tailored on patients' conditions and preferences [31-34] or in 1 study through telephone interviews via the interactive voice response (IVR) call self-management support program (ie, tailored messages, structured email) [35]. In addition, in 2 studies, tools were administered through simple telephone interviews [36,37]. In 1 study, questionnaires were delivered through tablets available from a site to assess the psychological well-being of patients [38], whereas another study used a mobile app [39]. Furthermore, 3 studies were conducted using the mood rating scale, which consists of asking the patient "how do you feel?" through a tailored text message–based diabetes support system (ie, tablets, mobile phone) [38,40,41]. Specifically, the mood rating scale is integrated into the technology to evaluate moods and their patterns over time. The research question concerning the efficacy and usability of the tools has not found a clear and complete answer. It seems that this research question was not fully addressed in previous research but just *tangentially* touched on or inferred as a parallel finding.

Advantages and Disadvantages of Technological Devices

The use of information and communication technology (ICT) apps in health care settings is increasing globally. Indeed, a useful way of communicating preventive methods to the population is through ICT [42,43]. This is motivated by an interest in facilitating active participation for people to self-manage their health as well as by the need to develop apps and platforms, which can be more cost-effective, compared with traditional approaches, and also to manage chronic conditions, such as diabetes [44,45]. For instance, the ATA or automated telephone communication systems is an app used to deliver both preventive health care programs and services to manage chronic conditions. Several studies have analyzed the ATA system for the management of diabetes [46,47], heart failure [48,49], coronary heart disease [50], and asthma [51] as well as for health-promoting methods, including dietary behavior [52,53] and physical activity [54,55]. ATA can deliver voice messages and gather health-related information from patients using voice recognition programs or touch-tone telephone [56], in addition to, or instead of, the telephone interaction between health professionals and patients. In particular, ATA has 3 subcategories: (1) unidirectional ATA, which enables one-way, noninteractive voice communication, including, for instance, interventions such as automated reminder calls to take medication; (2) the IVR system, which is the most common form of two-way real-time communication, allowing automated tailored feedback based on the monitoring of an individual's progress, thereby allowing one-to-one interventions [56,57]; and (3) ATA with additional functions, namely ATA Plus, such as access to an expert to request support and ask questions via telephone or face-to-face meetings, and also the delivery of automated, nonvoice communications such as SMS text messages or email [58]. ATA—conceived as a data collection tool—presents several advantages compared with the classical face-to-face assessment [59], such as simplicity, anonymity, and low costs [56,60]. Therefore, ATA can allow access to health care systems 7 days a week for 24 hours a day, together with immediate feedback to the patient [61,62]. Indeed, studies reported higher levels of user satisfaction experience, suggesting that it is accessible for both patients and health care professionals [62]. Unlike face-to-face interactions, which can evoke socially acceptable answers, leading to the underreporting of stigmatizing behaviors and overreporting of socially desirable behaviors, ATA can elicit better self-reporting of specific and sensitive problems (eg, alcohol and substance use) and reduce self-reporting bias [63] as well as health care delivery costs [64,65]. On the other hand, ATA may also have disadvantages such as the difficulty in catching, interpreting, or responding to patients' nonverbal answers to the interview questions [63,66]. Moreover, people with physical disabilities can have difficulty in using ATA [67], and some people can prefer face-to-face interaction rather than ATA [68]. In this framework, the classical telephone interviews, unlike the ATA system, can allow researchers and clinicians to obtain additional indications from the emphasis, intonation, hesitations, and the words used. However, as with the ATA system, they may have difficulty understanding the nonverbal responses. Telephone interviews also include limited telephone coverage in specific areas and lower response rates [69]. On the other hand, health professionals who use telephone interviews as assessment tools also have the opportunity to request a follow-up. Other digital solutions mentioned in this review are mobile phones, tablets, and desktop computers, in which mobile phone apps can allow real-time tracking of mood status, for instance, in patients with diabetes [70], anywhere and at any time, and they are particularly suitable for delivering immediate feedback, which makes them preferable to tablets and desktop computers. Moreover, besides being able to send voice and text messages, mobile phones present more advanced features, such as web searching, high-quality cameras, a GPS, and sound recording. Altogether with strong processors and operating systems, large memories, and high-resolution screens, mobile phones have turned into handheld computers. In particular, the use of mobile phones is increasing in health care settings (defined as mobile health [mHealth]), allowing health professionals to provide easy and rapid access to updated medical information [71,72]. Indeed, a great number of mHealth apps have become useful tools for health professionals, including health record maintenance and access, clinical software apps for suggestions within disease diagnosis, patient management and monitoring, clinical decision making, and medical training [71,73,74]. Moreover, mHealth has been found to support better clinical decision making and improve patient health outcomes [74,75]. On the other hand, one major issue concerns the security of health information delivered via mobile phones. mHealth adopts wireless atmospheric media to transmit data in the form of radio signals, which seem to be vulnerable to hackers and therefore to modification or distortion [76]; moreover, mHealth is closely networked with other wireless devices [77]. However, most professionals think that mHealth could significantly improve health care delivery processes, thereby improving patients' psychophysical health. Indeed, mHealth interventions can reduce costs, save time, facilitate access to medical information, and provide a simpler and quicker way for patients and clinicians to send medical communications. Therefore, the adoption of mHealth improves the lifestyle, nutrition, behaviors, and quality of life of people with various diseases, particularly with chronic conditions. Thus, mHealth is increasingly considered to be one of the best digital solutions for the support of individuals in the management of their disease and the improvement of their health conditions [77].

Principal Instruments to Assess Psychological Symptoms in T2DM in the Technology Field

Most studies included in the review reported only the name of the instrument, thereby not providing information regarding the psychometric properties of the tools, such as reliability, validity, data monitoring, contextual collection, and/or whether the instrument was specified for the population with diabetes [28,31,33-36,38,40]. In addition, the reasons for using such tools were not described in the analyzed papers; thus, it was not possible to ensure that the scales were used to address diabetes. In contrast, studies that administered the Diabetes Distress Scale-17 (DDS-17) [3] to evaluate diabetes-specific distress in the management of the disease reported that this tool shows good reliability and internal validity of the measure across independent samples [28,33,36,38]. This is the only scale that has been validated specifically for diabetes distress, focusing on problems that those patients may experience, such as emotional burden, interpersonal distress, or regimen-related distress [3].

Another instrument that is widely used to assess depressive symptoms through technologies is the Patient Health Questionnaire-9 (PHQ-9); indeed, 9 studies administered the above questionnaire [25-28,31-34,38]. The Centers for Medicare and Medicaid Services recommend the use of the PHQ-9 for home health care patients. The PHQ-9 was tested in primary care, demonstrating clinical relevance in relation to the Diagnostic and Statistical Manual of Mental Disorder-IV-Text Revision [24,78]. The PHQ-9 further comprises 2 components: symptoms and functional impairment assessment, which is useful for diagnosis, and a severity score, which is useful for selecting and monitoring treatments [26,27,32,38]. In addition, 1 study that administered the PHQ-8, a standardized and validated scale, showed its good reliability in assessing depressive symptoms [37]. Therefore, the PHQ represents a good tool to assess depressive symptoms in chronic diseases, including T2DM, as shown by its wide use. With regard to other psychological symptoms examined in the included studies, the tools used to assess anxiety symptoms were few, although the literature highlighted how anxiety could influence the chronic disease itself [79] to a higher degree than depressive disorders. In this context, 1 study administered the Hospital Anxiety and Depression Scale-14 items [80] to assess anxiety and depression symptoms [40], and 2 studies [30,35] evaluated depression symptoms using the Center for Epidemiological Studies Depression Scale [81]. Furthermore, 2 other studies [31,33] analyzed depression symptoms using the Hopkins Symptom Checklist Depression-20 [82]. With regard to the evaluation of the emotional distress experienced by people with diabetes, 2 studies [27,35] administered the Problem Areas in Diabetes [83]. Furthermore, 2 studies administered the Brief Symptom Inventory [84] to assess a wide range of psychological symptoms [33,36]. Stress symptoms were evaluated using a psychological scale through the mood rating; this was found to be interesting, as this scale is a technological modality to assess the nature of mood, thereby giving special importance to the monitoring instead of the screening of these symptoms [38,40,41]. The most used self-reports through technologies evaluating diabetes quality of life can be grouped into 3 categories: one refers to pain that can interfere with normal work and the other 2 categories refer to the physical and emotional symptoms related to the quality of life as assessed through the Medical Outcome Study Short-Form (MOS-SF) Health Survey 12 and 36 items [85,86], respectively, which are considered to be reliable and valid scales [26]. Overall, 6 studies administered the short version of the MOS [30,31,34,35,39,40] and 2 studies administered the longer version [25,38]. One study [26] evaluated the self-perception of quality of life through the EuroQol-5 Dimension [87], which has been tested and validated to capture the difference in the quality of life in patients with chronic diseases [28]. One study [29] analyzed the healthy self-management of the disease itself through the Health Education Impact Questionnaire [88], and it has also been validated in a primary health care context with patients with several chronic conditions, including diabetes [29]. Therefore, it seems that these studies used tools specifically targeted to patients with chronic diseases, including diabetes, to evaluate their quality of life as they had to change their physical and psychological habits. It is worth noting that one of the variables that emerged after the revision of the included studies was self-efficacy. Self-efficacy is a key variable in the proper management of diabetes within the health care setting [89]. Indeed, it represents the awareness and the perception that each individual has of their capacity to produce the desired results necessary to influence events, thereby affecting their lives [30]. Indeed, in the context of T2DM diabetes, a study found that higher glycemic control is associated with better self-efficacy and self-care behaviors [90].

A summary of the types of technological solutions through which patient screening and psychological assessment were conducted is presented in Multimedia Appendix 1 [25-41]. In addition, Textbox 1 shows the advantages and disadvantages of administering questionnaires in paper-and-pencil form versus those in digital form. Furthermore, all the questionnaires used to assess psychological symptoms in patients with T2DM are described in-depth in Multimedia Appendix 2 [3,79,82-90].

Textbox 1. Advantages and disadvantages of administering paper-and-pencil questionnaires versus administering questionnaires through digital solutions.

Paper-and-pencil questionnaires

- Advantages
    - More cost-effective when surveying data in small samples
    - Could increase a working alliance because of human interaction
    - Benefits people who do not have internet access
    - More favorable format when it comes to longer questionnaires
    - No digital skills are required to answer
- Disadvantages
    - Lack of immediate data analysis
    - Data entry is needed to store them in databases
    - Handwritten responses could be difficult to interpret, especially when it comes to open-ended questions
    - Needs longer data processing
    - Printing and archiving of the questionnaires are needed
    - There is the option of skipping questions
    - Human errors can occur when updating the database

Questionnaires through digital solutions

- Advantages
    - Save time for clinicians and researchers
    - Faster in delivering the questionnaires
    - Directly collect data on the web
    - Furnish an immediate feedback
    - Collect data from people around the world
    - More ecological (ie, no printing and other costs at the point of completing setup)
    - Better graphic layout (ie, not only color images and text, but also dynamic and interactive animation)
    - Collect all the responses, thereby avoiding unanswered questions
    - Allows longer answers to open-ended questions
    - Gives more time to fill in the questionnaires
- Disadvantages
    - Familiarity with digital devices is required
    - Possible difficulties for data analysis derived from people filling-in the questionnaire multiple times, which would bias the results
    - Lack of technological devices
    - Technical problems
    - Unreliable network

Discussion

Principal Findings

This study aimed to review past literature regarding the types of technological devices used to administer psychological instruments for assessing psychological distress (ie, depressive, anxiety symptoms) and stress in patients with T2DM. Assessing diabetes-related psychological symptoms can be challenging, due in part to the complexity of diabetes care. Indeed, diabetes self-care should be multidimensional, including treatment for both organic and psychological symptoms, whereas at the same time, it should use technology-based tools, which show good psychometric properties and are validated by samples of patients with chronic diseases.

First, studies seem to focus on the evaluation of psychological distress and diabetes-specific distress, particularly anxiety, depression, and stress symptoms. Indeed, instruments that measure symptoms of anxiety, stress, and depression related to

T2DM, administered through technologies, were analyzed, and their efficacy and usability were evaluated. In this context, the psychometric properties of the instruments are prerequisites for an accurate assessment of psychological distress in patients with diabetes. Inadequate reliability and validity of the tools make it difficult to detect the psychological well-being of patients with diabetes and the impact of interventions on their well-being or quality of life. Here, the timing of the test can influence its reliability and validity, and therefore it needs to be taken into consideration. Within a longitudinal evaluation of an intervention, if the duration of the test was too short, participants could recall information from the first time they completed it, which could bias the findings. Alternatively, if its duration was too long, participants may have changed significantly, which could also bias the results [91].

In most studies, screening was conducted using written questionnaires at baseline and after a follow-up as well as telephone interviews (ie, simple telephone interviews, ATA calls, and IVR) to assess psychological symptoms [25-32,36,37].

Few studies have used digital solutions, such as mobile apps, tablets, and computers, to deliver psychological self-reports for intervention groups, even though they were investigating the psychological symptoms related to the disease itself [26,31,33,35-39,84]. In this context, the recent progress in technologies supports the ecological momentary assessment of mood, using mood ratings through mobile devices, outside the clinical environment. The mood rating scale can help to bypass issues associated with infrequent reporting of depressive symptoms and allows for a better representation of the dynamic nature of mood, which is often left unreported, and to better guide treatment planning [92,93]. For instance, delivering psychological instruments through technologies (eg, mobile phone apps) allows researchers to collect data directly from the web; thus, the wide use of the mood rating strategy would suggest saving time [94,95]. For instance, some studies have analyzed the feasibility of daily or weekly SMS text messages based on mood ratings, showing that mood ratings represent a valid monitoring strategy for patients with depression [70,96,97]. Mobile phone apps allow real-time tracking of mood status in patients with T2DM [70], anywhere and at any time, and they are particularly suitable for delivering immediate feedback (which makes them preferable to tablets and desktop computers). Moreover, they facilitate data collection in a more contextualized, pervasive, longitudinal, and reliable way, rather than using written questionnaires (preintervention and postintervention). In addition, they allow adapting the intervention to patients' needs, supporting them in the management of their chronic disease. On the basis of other chronic diseases, mobile phone apps can provide psychoeducation [98], smoking cessation support [99], cognitive behavioral therapy [100], and support to caregivers [101]. Therefore, the identification of the appropriate psychological tools that can be embedded in digital devices, such as smartphones, could allow researchers and clinicians to conduct a screening of the level of anxiety, stress, and depressive symptoms in patients with T2DM in a faster and easier way.

Second, the PHQ-9 and the DDS-17 emerged as useful tools for the assessment of depressive symptoms in chronic diseases and for problems often experienced by these patients, such as emotional burden, interpersonal distress, or regimen-related distress [3]. Another widely used instrument is the MOS-SF-12, which assesses the physical and emotional symptoms related to the quality of life.

Moreover, considering the importance of the role of self-efficacy in the management of emotion-related diabetes, it is recommended that researchers and clinicians use specific tools to address self-efficacy in those patients. Of particular note is the fact that most studies used psychological tools in the standard paper format to assess the effectiveness of interventions without integrating them into digital solutions. Indeed, in this review, few studies delivered web-based questionnaires [38,39]. As an example, the Meru Health Ascend, a smartphone-based, therapist-supported intervention for depression and anxiety in patients with no chronic disease, delivered 2 validated scales for psychological distress (ie, PHQ-9 and the Generalized Anxiety Disorder-7) on a smartphone [102].

In this study, the integration of assessment instruments in digital solutions improved the assessment of depressive and anxiety symptoms. Thus, research suggests that mHealth interventions are functional ways of supporting the treatment of depression and anxiety symptoms [102].

Limitations and Strengths

This review presents some limitations, as it included only papers in English, which limits the generalization of the findings. In addition, a limitation could be identified in the different implications regarding the use of technological devices from the age of 18 to 70 years. Older adults may be less familiar with the use of these devices. Future work should evaluate the effect of the use of technological devices among individuals of different ages. In the context of digital solutions, the disadvantage of smartphone-, tablet-, or computer-based apps is that they can be removed by the user; however, such device-based apps and other conversational agents can represent a valuable solution in administering psychological tools for the screening of patients, especially in emergency situations such as the SARS-CoV-2 pandemic. Furthermore, most studies do not provide information regarding the psychometric properties of tools, such as reliability, validity, and contextual collection. It could be important to include such characteristics to better understand the instruments used for assessing psychological distress and diabetes-specific distress. Another limitation is the noninclusion of videoconference calling as a digital solution, which especially during the COVID-19 lockdown represented the method of choice among practitioners in psychology and psychiatry for a remote assessment of mental health. Future studies should include these types of technological solutions to expand this literature review.

Finally, the focus of the review was only on psychological measures related to T2DM. In future works, one would recommend the integration of the assessment and monitoring of both organic and psychological symptoms in patients with T2DM.

On the other hand, this review also has some strengths. In addition to depression, the review considered all possible psychological symptoms related to T2DM.

Future Development and Implications of the Study

This review highlighted implications that may have an impact on future research and clinical practice. In particular, the use of appropriate technological solutions to assess the psychological condition of patients can allow early detection of depression, stress, and anxiety symptoms, especially in chronic conditions as well as mental disorders [103]. In less severe cases, it might also help the deployment of mHealth interventions to support and improve depression, anxiety, and stress symptoms in T2DM to lower the burden for the health care system, in a stepped care approach [104]. Previous research found that mHealth solutions were well accepted by young adults [105], and further research is required to assess their acceptability by older adults with T2DM. Furthermore, because of the huge growth of mHealth apps, it would be useful to screen the levels of anxiety, stress, and depression symptoms in patients with diabetes, using appropriate psychological instruments. This screening could be able to identify whether the patient showed mild, moderate, or severe symptoms, thereby allowing clinicians to better set up interventions.

Conclusions

In view of the large increase in the number of patients with diabetes globally, it is important to intensify efforts in the deployment of digital solutions for the accurate assessment of patients' psychological condition. These instruments can be useful for clinicians and researchers to better monitor patients' conditions. For example, the mood rating scale is widely used in studies to assess diabetes-related stress and depression. Therefore, the identification of appropriate psychological tools that can be deployed through mHealth solutions allows researchers and clinicians to screen for anxiety, stress, and depression in patients with T2DM in a faster and easier way. Moreover, data from the literature suggest that mHealth interventions are preferable to other types of digital interventions [101]. For instance, mHealth apps have recently been deployed in the field of psychology to deliver evidence-based treatments for depression and anxiety and to overcome barriers of face-to-face psychotherapy [101]. Therefore, this study can add to the scientific body of literature on the revision of valuable digital solutions in assessing and monitoring the mental health of patients, in which the traditional paper-and-pencil instruments can be delivered through digital solutions. Indeed, administering questionnaires through technological devices is a more cost-effective, time-efficient method for data gathering and for real-time tracking of mood status. Effective programs for chronic disease management should combine relevant information systems, with constant follow-up and targeted self-management for patients. In this way, ICT is incorporated in such a way as to provide accessible and convenient psychoeducational information as well as self‐management tools for people with long-term conditions. Finally, it is worth noting that ICT represents the only feasible way to assist and maintain health care, in chronic and nonchronic disease, in such a dramatic period as the actual pandemic, and thus gaining better knowledge and flexibility in the use of these methods becomes essential.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Articles from the literature review.
[DOCX File , 25 KB - jmir_v23i1e17740_app1.docx]

Multimedia Appendix 2
Description of the psychological instruments.
[DOCX File , 22 KB - jmir_v23i1e17740_app2.docx]

References

1. Egede LE, Zheng D, Simpson K. Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes. Diabetes Care 2002 Mar;25(3):464-470. [doi: 10.2337/diacare.25.3.464] [Medline: 11874931]
2. Gonzalez J, Esbitt S, Schneider H, Osborne P, Kupperman E. Psychological Comorbidities of Type 2 Diabetes. New York, USA: Routedge; 2011:-122.
3. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. Diabetes Care 2005 Mar;28(3):626-631. [doi: 10.2337/diacare.28.3.626] [Medline: 15735199]
4. Young-Hyman D, de Groot M, Hill-Briggs F, Gonzalez J, Hood K, Peyrot M. Response to comments on young-hyman et al psychosocial care for people with diabetes: a position statement of the American diabetes association. Diabetes care 2016;39:2126-2140. Diabetes Care 2017 Sep;40(9):e131-e132 [FREE Full text] [doi: 10.2337/dci16-0051] [Medline: 28830965]
5. Ali S, Stone M, Skinner TC, Robertson N, Davies M, Khunti K. The association between depression and health-related quality of life in people with type 2 diabetes: a systematic literature review. Diabetes Metab Res Rev 2010 Feb;26(2):75-89. [doi: 10.1002/dmrr.1065] [Medline: 20186998]

6. de Groot M, Golden SH, Wagner J. Psychological conditions in adults with diabetes. Am Psychol 2016 Oct;71(7):552-562 [FREE Full text] [doi: 10.1037/a0040408] [Medline: 27690484]
7. Fisher L, Glasgow RE, Strycker LA. The relationship between diabetes distress and clinical depression with glycemic control among patients with type 2 diabetes. Diabetes Care 2010 May;33(5):1034-1036 [FREE Full text] [doi: 10.2337/dc09-2175] [Medline: 20150291]
8. Fisher L, Hessler DM, Polonsky WH, Mullan J. When is diabetes distress clinically meaningful?: establishing cut points for the Diabetes Distress Scale. Diabetes Care 2012 Feb;35(2):259-264 [FREE Full text] [doi: 10.2337/dc11-1572] [Medline: 22228744]
9. Winchester RJ, Williams JS, Wolfman TE, Egede LE. Depressive symptoms, serious psychological distress, diabetes distress and cardiovascular risk factor control in patients with type 2 diabetes. J Diabetes Complications 2016 Mar;30(2):312-317 [FREE Full text] [doi: 10.1016/j.jdiacomp.2015.11.010] [Medline: 26657725]
10. Nguyen H, Tran T, Nguyen C, Tran T, Tran B, Latkin C, et al. Impact of comorbid chronic conditions to quality of life among elderly patients with diabetes mellitus in Vietnam. Int J Environ Res Public Health 2019 Feb 13;16(4):- [FREE Full text] [doi: 10.3390/ijerph16040531] [Medline: 30781767]
11. Niti M, Ng T, Kua EH, Ho RC, Tan CH. Depression and chronic medical illnesses in Asian older adults: the role of subjective health and functional status. Int J Geriatr Psychiatry 2007 Nov;22(11):1087-1094. [doi: 10.1002/gps.1789] [Medline: 17407107]
12. Nguyen TT, Vu HT, Nguyen TN, Dao HT, Nguyen TX, Nguyen HT, et al. Assessment of nutritional status in older diabetic outpatients and related factors in Hanoi, Vietnam. JMDH 2019 Jul;Volume 12:601-606. [doi: 10.2147/jmdh.s194155]
13. Diabetes Distress. Diabetes. URL: https://www.diabetes.org.uk/professionals/resources/shared-practice/psychological-care/emotional-health-professionals-guide/chapter-3-diabetes-distress [accessed 2020-07-01]
14. Arvidsdotter T, Marklund B, Kylén S, Taft C, Ekman I. Understanding persons with psychological distress in primary health care. Scand J Caring Sci 2016 Dec;30(4):687-694. [doi: 10.1111/scs.12289] [Medline: 26463897]
15. Fisher L, Skaff MM, Mullan JT, Arean P, Glasgow R, Masharani U. A longitudinal study of affective and anxiety disorders, depressive affect and diabetes distress in adults with Type 2 diabetes. Diabet Med 2008 Sep;25(9):1096-1101 [FREE Full text] [doi: 10.1111/j.1464-5491.2008.02533.x] [Medline: 19183314]
16. Ansari M, Shams SH, Hakami M. The effect of training anger control and stress management on blood glucose levels in people with diabetes. HPJ In Persian 2011;1(3):47-60.
17. Amiri M, Aghaei A, Abedi A, Safari Y. Effectiveness of stress inoculation training (SIT) on depression Diabetics. HPJ In Persian 2014;3(1):16-28.
18. Mikaliūkštienė A, Žagminas K, Juozulynas A, Narkauskaitė L, Sąlyga J, Jankauskienė K, et al. Prevalence and determinants of anxiety and depression symptoms in patients with type 2 diabetes in Lithuania. Med Sci Monit 2014 Feb 4;20:182-190 [FREE Full text] [doi: 10.12659/MSM.890019] [Medline: 24492643]
19. Wagner JA, Tennen H, Finan PH, White WB, Burg MM, Ghuman N. Lifetime history of depression, type 2 diabetes, and endothelial reactivity to acute stress in postmenopausal women. Int J Behav Med 2012 Dec;19(4):503-511. [doi: 10.1007/s12529-011-9190-5] [Medline: 21964983]
20. de Groot M, Pinkerman B, Wagner J, Hockman E. Depression treatment and satisfaction in a multicultural sample of type 1 and type 2 diabetic patients. Diabetes Care 2006 Mar;29(3):549-553. [doi: 10.2337/diacare.29.03.06.dc05-1396] [Medline: 16505504]
21. Liberman A, Buckingham B, Phillip M. Diabetes technology and the human factor. Diabetes Technol Ther 2014 Feb;16(Suppl 1):S110-S118 [FREE Full text] [doi: 10.1089/dia.2014.1513] [Medline: 24479592]
22. Lu Y, Xu J, Zhao W, Han H. Measuring self-care in persons with type 2 diabetes: a systematic review. Eval Health Prof 2016 Jun;39(2):131-184 [FREE Full text] [doi: 10.1177/0163278715588927] [Medline: 26130465]
23. Cambridge Dictionary. URL: https://dictionary.cambridge.org/dictionary/english/device [accessed 2020-07-01]
24. Diagnostic and Statistical Manual of Mental Disorders. 4th Edition. Milano: Masson; 2000.
25. Quinn CC, Khokhar B, Weed K, Barr E, Gruber-Baldini AL. Older adult self-efficacy study of mobile phone diabetes management. Diabetes Technol Ther 2015 Jul;17(7):455-461 [FREE Full text] [doi: 10.1089/dia.2014.0341] [Medline: 25692373]
26. Islam SM, Lechner A, Ferrari U, Froeschl G, Alam DS, Holle R, et al. Mobile phone intervention for increasing adherence to treatment for type 2 diabetes in an urban area of Bangladesh: protocol for a randomized controlled trial. BMC Health Serv Res 2014 Nov 26;14:586 [FREE Full text] [doi: 10.1186/s12913-014-0586-1] [Medline: 25424425]
27. Welch G, Allen NA, Zagarins SE, Stamp KD, Bursell S, Kedziora RJ. Comprehensive diabetes management program for poorly controlled Hispanic type 2 patients at a community health center. Diabetes Educ 2011;37(5):680-688. [doi: 10.1177/0145721711416257] [Medline: 21918206]
28. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]

29. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mhealth app intervention for persons with type 2 diabetes and its associations with initial self-management: randomized controlled trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]
30. Weinstock RS, Brooks G, Palmas W, Morin PC, Teresi JA, Eimicke JP, et al. Lessened decline in physical activity and impairment of older adults with diabetes with telemedicine and pedometer use: results from the IDEATel study. Age Ageing 2011 Jan;40(1):98-105. [doi: 10.1093/ageing/afq147] [Medline: 21081539]
31. Ramirez M, Wu S, Jin H, Ell K, Gross-Schulman S, Myerchin Sklaroff L, et al. Automated remote monitoring of depression: acceptance among low-income patients in diabetes disease management. JMIR Ment Health 2016 Jan 25;3(1):e6 [FREE Full text] [doi: 10.2196/mental.4823] [Medline: 26810139]
32. Vidyanti I, Wu B, Wu S. Low-income minority patient engagement with automated telephonic depression assessment and impact on health outcomes. Qual Life Res 2015 May;24(5):1119-1129 [FREE Full text] [doi: 10.1007/s11136-014-0900-8] [Medline: 25543270]
33. Wu S, Ell K, Gross-Schulman SG, Sklaroff LM, Katon WJ, Nezu AM, et al. Technology-facilitated depression care management among predominantly Latino diabetes patients within a public safety net care system: comparative effectiveness trial design. Contemp Clin Trials 2014 Mar;37(2):342-354. [doi: 10.1016/j.cct.2013.11.002] [Medline: 24215775]
34. Hay JW, Lee P, Jin H, Guterman JJ, Gross-Schulman S, Ell K, et al. Cost-effectiveness of a technology-facilitated depression care management adoption model in safety-net primary care patients with type 2 diabetes. Value Health 2018 May;21(5):561-568 [FREE Full text] [doi: 10.1016/j.jval.2017.11.005] [Medline: 29753353]
35. Aikens JE, Rosland A, Piette JD. Improvements in illness self-management and psychological distress associated with telemonitoring support for adults with diabetes. Prim Care Diabetes 2015 Apr;9(2):127-134 [FREE Full text] [doi: 10.1016/j.pcd.2014.06.003] [Medline: 25065270]
36. Rotheram-Borus MJ, Tomlinson M, Gwegwe M, Comulada WS, Kaufman N, Keim M. Diabetes buddies: peer support through a mobile phone buddy system. Diabetes Educ 2012;38(3):357-365 [FREE Full text] [doi: 10.1177/0145721712444617] [Medline: 22546740]
37. Trief PM, Fisher L, Sandberg J, Cibula DA, Dimmock J, Hessler DM, et al. Health and psychosocial outcomes of a telephonic couples behavior change intervention in patients with poorly controlled type 2 diabetes: a randomized clinical trial. Diabetes Care 2016 Dec;39(12):2165-2173 [FREE Full text] [doi: 10.2337/dc16-0035] [Medline: 27456837]
38. Holland-Carter L, Tuerk PW, Wadden TA, Fujioka KN, Becker LE, Miller-Kovach K, et al. Impact on psychosocial outcomes of a nationally available weight management program tailored for individuals with type 2 diabetes: Results of a randomized controlled trial. J Diabetes Complications 2017 May;31(5):891-897 [FREE Full text] [doi: 10.1016/j.jdiacomp.2017.01.022] [Medline: 28319001]
39. Munster-Segev M, Fuerst O, Kaplan SA, Cahn A. Incorporation of a stress reducing mobile app in the care of patients with type 2 diabetes: a prospective study. JMIR Mhealth Uhealth 2017 May 29;5(5):e75 [FREE Full text] [doi: 10.2196/mhealth.7408] [Medline: 28554881]
40. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 5;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
41. Dobson R, Carter K, Cutfield R, Hulme A, Hulme R, McNamara C, et al. Diabetes text-message self-management support program (SMS4BG): a pilot study. JMIR Mhealth Uhealth 2015 Mar 25;3(1):e32 [FREE Full text] [doi: 10.2196/mhealth.3988] [Medline: 25830952]
42. Baranowski T, Frankel L. Gaming and technology for weight control and health promotion in children. Child Obesity 2012;8(1):7.
43. Haluza D, Jungwirth D. ICT and the future of health care: aspects of health promotion. Int J Med Inform 2015 Jan;84(1):48-57. [doi: 10.1016/j.ijmedinf.2014.09.005] [Medline: 25293532]
44. Carolan I, Smith T, Hall A, Swallow VM. Emerging communities of child-healthcare practice in the management of long-term conditions such as chronic kidney disease: qualitative study of parents' accounts. BMC Health Serv Res 2014 Jul 7;14:292 [FREE Full text] [doi: 10.1186/1472-6963-14-292] [Medline: 25001236]
45. Demain S, Gonçalves AC, Areia C, Oliveira R, Marcos AJ, Marques A, et al. Living with, managing and minimising treatment burden in long term conditions: a systematic review of qualitative research. PLoS One 2015;10(5):e0125457 [FREE Full text] [doi: 10.1371/journal.pone.0125457] [Medline: 26024379]
46. Katalenich B, Shi L, Liu S, Shao H, McDuffie R, Carpio G, et al. Evaluation of a remote monitoring system for diabetes control. Clin Ther 2015 Jun 1;37(6):1216-1225 [FREE Full text] [doi: 10.1016/j.clinthera.2015.03.022] [Medline: 25869625]
47. Khanna R, Stoddard PJ, Gonzales EN, Villagran-Flores M, Thomson J, Bayard P, et al. An automated telephone nutrition support system for Spanish-speaking patients with diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1115-1120 [FREE Full text] [doi: 10.1177/1932296814550186] [Medline: 25239122]
48. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309. [doi: 10.1056/nejmoa1010029]

49. Krum H, Forbes A, Yallop J, Driscoll A, Croucher J, Chan B, et al. Telephone support to rural and remote patients with heart failure: the Chronic Heart Failure Assessment by Telephone (CHAT) study. Cardiovasc Ther 2013 Aug;31(4):230-237 [FREE Full text] [doi: 10.1111/1755-5922.12009] [Medline: 23061492]
50. Reid RD, Pipe AL, Quinlan B, Oda J. Interactive voice response telephony to promote smoking cessation in patients with heart disease: a pilot study. Patient Educ Couns 2007 Jun;66(3):319-326. [doi: 10.1016/j.pec.2007.01.005] [Medline: 17336026]
51. Bender BG, Apter A, Bogen DK, Dickinson P, Fisher L, Wamboldt FS, et al. Test of an interactive voice response intervention to improve adherence to controller medications in adults with asthma. J Am Board Fam Med 2010;23(2):159-165 [FREE Full text] [doi: 10.3122/jabfm.2010.02.090112] [Medline: 20207925]
52. Delichatsios HK, Friedman RH, Glanz K, Tennstedt S, Smigelski C, Pinto BM, et al. Randomized trial of a 'talking computer' to improve adults' eating habits. Am J Health Promot 2001;15(4):215-224. [doi: 10.4278/0890-1171-15.4.215] [Medline: 11349340]
53. Wright JA, Phillips B, Watson B, Newby P, Norman G, Adams W. Randomized trial of a family-based, automated, conversational obesity treatment program for underserved populations. Obesity (Silver Spring) 2013 Sep;21(9):E369-E378 [FREE Full text] [doi: 10.1002/oby.20388] [Medline: 23512915]
54. David P, Buckworth J, Pennell ML, Katz ML, DeGraffinreid CR, Paskett ED. A walking intervention for postmenopausal women using mobile phones and Interactive Voice Response. J Telemed Telecare 2012 Jan;18(1):20-25 [FREE Full text] [doi: 10.1258/jtt.2011.110311] [Medline: 22052963]
55. Pinto BM, Friedman R, Marcus BH, Kelley H, Tennstedt S, Gillman MW. Effects of a computer-based, telephone-counseling system on physical activity. Am J Prev Med 2002 Aug;23(2):113-120. [doi: 10.1016/s0749-3797(02)00441-5]
56. Piette J, Lun K, Moura L, Fraser H, Mechael P, Powell J, et al. Impacts of e-health on the outcomes of care in low- and middle-income countries: where do we go from here? Bull World Health Org 2012 May 1;90(5):365-372. [doi: 10.2471/blt.11.099069]
57. Rosen RC, Stephens-Shields AJ, Cunningham GR, Cifelli D, Cella D, Farrar JT, et al. Comparison of interactive voice response (IVR) with paper administration of instruments to assess functional status, sexual function, and quality of life in elderly men. Qual Life Res 2016 Apr;25(4):811-821 [FREE Full text] [doi: 10.1007/s11136-015-1133-1] [Medline: 26358063]
58. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
59. Lee H, Friedman ME, Cukor P, Ahern D. Interactive voice response system (IVRS) in health care services. Nursing Outlook 2003 Dec;51(6):277-283. [doi: 10.1016/s0029-6554(03)00161-1]
60. Hall JA, Huber DL. Telephone management in substance abuse treatment. Telemed J E Health 2000;6(4):401-407. [doi: 10.1089/15305620050503870] [Medline: 11242548]
61. Abu-Hasaballah K, James A, Aseltine RH. Lessons and pitfalls of interactive voice response in medical research. Contemp Clin Trials 2007 Sep;28(5):593-602. [doi: 10.1016/j.cct.2007.02.007] [Medline: 17400520]
62. Schroder KE, Johnson CJ. Interactive voice response technology to measure HIV-related behavior. Curr HIV/AIDS Rep 2009 Nov;6(4):210-216. [doi: 10.1007/s11904-009-0028-6] [Medline: 19849964]
63. Phillips L, Hendren S, Humiston S, Winters P, Fiscella K. Improving breast and colon cancer screening rates: a comparison of letters, automated phone calls, or both. J Am Board Fam Med 2015;28(1):46-54 [FREE Full text] [doi: 10.3122/jabfm.2015.01.140174] [Medline: 25567822]
64. Szilagyi PG, Albertin C, Humiston SG, Rand CM, Schaffer S, Brill H, et al. A randomized trial of the effect of centralized reminder/recall on immunizations and preventive care visits for adolescents. Acad Pediatr 2013;13(3):204-213 [FREE Full text] [doi: 10.1016/j.acap.2013.01.002] [Medline: 23510607]
65. Williams C, Whitfield G. Written and computer-based self-help treatments for depression. Br Med Bull 2001;57:133-144. [doi: 10.1093/bmb/57.1.133] [Medline: 11719913]
66. Mundt J, Ferber K, Rizzo M, Greist J. Computer-automated dementia screening using a touch-tone telephone. Arch Intern Med 2001 Nov 12;161(20):2481-2487. [doi: 10.1001/archinte.161.20.2481] [Medline: 11700161]
67. Mahoney DF, Tarlow BJ, Jones RN. Effects of an automated telephone support system on caregiver burden and anxiety: findings from the REACH for TLC intervention study. Gerontologist 2003 Aug;43(4):556-567. [doi: 10.1093/geront/43.4.556] [Medline: 12937334]
68. Groves RM. Theories and methods of telephone surveys. Annu Rev Sociol 1990 Aug;16(1):221-240. [doi: 10.1146/annurev.so.16.080190.001253]
69. Zhang MW, Ward J, Ying JJ, Pan F, Ho RC. The alcohol tracker application: an initial evaluation of user preferences. BMJ Innov 2016 Jan;2(1):8-13 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000087] [Medline: 27019744]
70. Wallace S, Clark M, White J. 'It's on my iPhone': attitudes to the use of mobile computing devices in medical education, a mixed-methods study. BMJ Open 2012;2(4):- [FREE Full text] [doi: 10.1136/bmjopen-2012-001099] [Medline: 22923627]

71. Jamal A, Temsah M, Khan SA, Al-Eyadhy A, Koppel C, Chiang MF. Mobile phone use among medical residents: a cross-sectional multicenter survey in Saudi Arabia. JMIR Mhealth Uhealth 2016 May 19;4(2):e61 [FREE Full text] [doi: 10.2196/mhealth.4904] [Medline: 27197618]
72. O?Neill K, Holmer H, Greenberg S, Meara J. Applying surgical appsmartphone and tablet apps prove useful in clinical practice. Bull Am Coll Surg 2013;98(11):18.
73. Divall P, Camosso-Stefinovic J, Baker R. The use of personal digital assistants in clinical decision making by health care professionals: a systematic review. Health Informatics J 2013 Mar;19(1):16-28 [FREE Full text] [doi: 10.1177/1460458212446761] [Medline: 23486823]
74. Aungst TD. Medical applications for pharmacists using mobile devices. Ann Pharmacother 2013 Jul 2;47(7-8):1088-1095. [doi: 10.1345/aph.1s035]
75. Bajwa M. Wireless Network Security Threats and Mitigation—A Survey. Open Journal of Business and Management. 2014. URL: https://file.scirp.org/Html/4-1530021_49815.htm [accessed 2020-12-11]
76. Choi CQ. Digital danger. Sci Am 2012 Nov 13;307(6):14-14. [doi: 10.1038/scientificamerican1212-14]
77. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
78. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
79. Collins M, Corcoran P, Perry I. Anxiety and depression symptoms in patients with diabetes. Diabet Med 2009 Feb;26(2):153-161. [doi: 10.1111/j.1464-5491.2008.02648.x] [Medline: 19236618]
80. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
81. Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. J Aging Health 1993 May;5(2):179-193. [doi: 10.1177/089826439300500202] [Medline: 10125443]
82. Derogatis L, Lipman R, Rickels K, Uhlenhuth E, Covi L. The hopkins symptom checklist (HSCL). A measure of primary symptom dimensions. Mod Probl Pharmacopsychiatry 1974;7:79-110. [doi: 10.1159/000395070] [Medline: 4607278]
83. Hermanns N, Kulzer B, Krichbaum M, Kubiak T, Haak T. How to screen for depression and emotional problems in patients with diabetes: comparison of screening characteristics of depression questionnaires, measurement of diabetes-specific emotional problems and standard clinical assessment. Diabetologia 2006 Mar;49(3):469-477. [doi: 10.1007/s00125-005-0094-2] [Medline: 16432706]
84. Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. Psychol Med 2009 Jul 9;13(3):595-605. [doi: 10.1017/s0033291700048017]
85. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
86. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). Medical Care 1992;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]
87. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
88. Osborne RH, Elsworth GR, Whitfield K. The health education impact questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
89. Bandura A. Self-Efficacy. The Exercise of Control. New York, USA: Freeman; 1994.
90. D'Souza MS, Karkada SN, Parahoo K, Venkatesaperumal R, Achora S, Cayaban AR. Self-efficacy and self-care behaviours among adults with type 2 diabetes. Appl Nurs Res 2017 Aug;36:25-32. [doi: 10.1016/j.apnr.2017.05.004] [Medline: 28720235]
91. McLeod S. What is Reliability? Simple Psychology. 2007. URL: https://www.simplypsychology.org/reliability.html [accessed 2020-01-01]
92. Wenze SJ, Miller IW. Use of ecological momentary assessment in mood disorders research. Clin Psychol Rev 2010 Aug;30(6):794-804. [doi: 10.1016/j.cpr.2010.06.007] [Medline: 20619520]
93. Aguilera A, Schueller SM, Leykin Y. Daily mood ratings via text message as a proxy for clinic based depression assessment. J Affect Disord 2015 Apr 1;175:471-474 [FREE Full text] [doi: 10.1016/j.jad.2015.01.033] [Medline: 25679202]
94. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
95. Richmond SJ, Keding A, Hover M, Gabe R, Cross B, Torgerson D, et al. Feasibility, acceptability and validity of SMS text messaging for measuring change in depression during a randomised controlled trial. BMC Psychiatry 2015 Apr 3;15:68 [FREE Full text] [doi: 10.1186/s12888-015-0456-3] [Medline: 25886377]

96. Keding A, Böhnke JR, Croudace TJ, Richmond SJ, MacPherson H. Validity of single item responses to short message service texts to monitor depression: an mHealth sub-study of the UK ACUDep trial. BMC Med Res Methodol 2015 Jul 30;15:56 [FREE Full text] [doi: 10.1186/s12874-015-0054-6] [Medline: 26224088]
97. Fealy S, Chan S, Wynne O, Dowse E, Ebert L, Ho R, et al. The Support for New Mums Project: a protocol for a pilot randomized controlled trial designed to test a postnatal psychoeducation smartphone application. J Adv Nurs 2019 Jun;75(6):1347-1359. [doi: 10.1111/jan.13971] [Medline: 30740767]
98. Tran B, Le X, Nguyen P, Le Q, Mai H, Nguyen H, et al. Feasibility of e-health interventions on smoking cessation among Vietnamese active internet users. Int J Environ Res Public Health 2018 Jan 20;15(1):- [FREE Full text] [doi: 10.3390/ijerph15010165] [Medline: 29361694]
99. Zhang MW, Ho RC, Cassin SE, Hawa R, Sockalingam S. Online and smartphone based cognitive behavioral therapy for bariatric surgery patients: initial pilot study. THC 2015 Oct 27;23(6):737-744. [doi: 10.3233/thc-151026]
100. Zhang MW, Chan S, Wynne O, Jeong S, Hunter S, Wilson A, et al. Conceptualization of an evidence-based smartphone innovation for caregivers and persons living with dementia. THC 2016 Sep 14;24(5):769-773. [doi: 10.3233/thc-161165]
101. Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, et al. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. JMIR Mhealth Uhealth 2019 Aug 26;7(8):e14284 [FREE Full text] [doi: 10.2196/14284] [Medline: 31452521]
102. Zhang MW, Ho RC, Loh A, Wing T, Wynne O, Chan SW, et al. Current status of postnatal depression smartphone applications available on application stores: an information quality analysis. BMJ Open 2017 Nov 14;7(11):e015655 [FREE Full text] [doi: 10.1136/bmjopen-2016-015655] [Medline: 29138195]
103. Mayora O. The MONARCA project for bipolar disorder treatment. J CyberTherapy Rehab 2011:-. [doi: 10.1037/e573152011-003]
104. von Korff M, Tiemens B. Individualized stepped care of chronic illness. West J Med 2000 Feb;172(2):133-137 [FREE Full text] [doi: 10.1136/ewjm.172.2.133] [Medline: 10693379]
105. Do TT, Le MD, Van Nguyen T, Tran BX, Le HT, Nguyen HD, et al. Receptiveness and preferences of health-related smartphone applications among Vietnamese youth and young adults. BMC Public Health 2018 Jun 19;18(1):764 [FREE Full text] [doi: 10.1186/s12889-018-5641-0] [Medline: 29921258]

Abbreviations

ATA: automated telephonic assessment
DDS-17: Diabetes Distress Scale-17
ICT: information and communication technology
IVR: interactive voice response
mHealth: mobile health
MOS-SF: Medical Outcome Study Short-Form Health Survey
PHQ-9: Patient Health Questionnaire-9
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 09.01.20; peer-reviewed by D Di Riso, R Ho, R Barak Ventura; comments to author 10.03.20; revised version received 05.08.20; accepted 11.11.20; published 07.01.21.

Please cite as:
Bassi G, Gabrielli S, Donisi V, Carbone S, Forti S, Salcuni S
Assessment of Psychological Distress in Adults With Type 2 Diabetes Mellitus Through Technologies: Literature Review
J Med Internet Res 2021;23(1):e17740
URL: https://www.jmir.org/2021/1/e17740
doi:10.2196/17740
PMID:33410762

Review

Effect of Interactive eHealth Interventions on Improving Medication Adherence in Adults With Long-Term Medication: Systematic Review

Bart P H Pouls[1,2], PharmD; Johanna E Vriezekolk[1], PhD; Charlotte L Bekker[2], PhD; Annemiek J Linn[3], PhD; Hein A W van Onzenoort[4,5], PharmD, PhD; Marcia Vervloet[6], PhD; Sandra van Dulmen[6,7,8], Prof Dr; Bart J F van den Bemt[1,2], PharmD, Prof Dr

[1]Department of Rheumatology Research, Sint Maartenskliniek, Nijmegen, Netherlands
[2]Department of Pharmacy, Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, Netherlands
[3]Amsterdam School of Communication Research, University of Amsterdam,, Amsterdam, Netherlands
[4]Department of Clinical Pharmacy, Amphia Hospital, Breda, Netherlands
[5]Department of Clinical Pharmacy and Toxicology, Maastricht University Medical Centre, Maastricht, Netherlands
[6]Nivel (Netherlands Institute for Health Services Research), Utrecht, Netherlands
[7]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, Netherlands
[8]Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway

Corresponding Author:
Bart P H Pouls, PharmD
Department of Rheumatology Research
Sint Maartenskliniek
Hengstdal 3
Nijmegen, 6574NA
Netherlands
Phone: 31 0243272518
Email: b.pouls@maartenskliniek.nl

Abstract

Background: Medication nonadherence leads to suboptimal treatment outcomes, making it a major priority in health care. eHealth provides an opportunity to offer medication adherence interventions with minimal effort from health care providers whose time and resources are limited.

Objective: The aim of this systematic review is twofold: (1) to evaluate effectiveness of recently developed and tested interactive eHealth (including mHealth) interventions on medication adherence in adult patients using long-term medication and (2) to describe strategies among effective interventions.

Methods: MEDLINE, EMBASE, Cochrane Library, PsycINFO, and Web of Science were systematically searched from January 2014 to July 2019 as well as reference lists and citations of included articles. Eligible studies fulfilled the following inclusion criteria: (1) randomized controlled trial with a usual care control group; (2) a total sample size of at least 50 adult patients using long-term medication; (3) applying an interactive eHealth intervention aimed at the patient or patient's caregiver; and (4) medication adherence as primary outcome. Methodologic quality was assessed using the Cochrane risk of bias tool. Selection and quality assessment of studies were performed by 2 researchers (BP and BvdB or JV) independently. A best evidence synthesis was performed according to the Cochrane Back Review Group.

Results: Of the 9047 records screened, 22 randomized clinical trials were included reporting on 29 interventions. Most (21/29, 72%) interventions specified using a (mobile) phone for calling, SMS text messaging, or mobile apps. A majority of all interactive interventions (17/29) had a statistically significant effect on medication adherence ($P<.05$). Of these interventions, 9 had at least a small effect size (Cohen $d \geq 0.2$) and 3 showed strong odds for becoming adherent in the intervention group (odds ratio > 2.0). Our best evidence synthesis provided strong evidence for a positive effect of interventions using SMS text messages or interactive voice response, mobile app, and calls as mode of providing adherence tele-feedback. Intervention strategies "to teach medication management skills," "to improve health care quality by coordinating medication adherence care between professionals," and "to facilitate communication or decision making between patients and health care providers" also showed strong evidence for a positive effect.

Conclusions: Overall, this review supports the hypothesis that interactive eHealth interventions can be effective in improving medication adherence. Intervention strategies that improve patients' treatment involvement and their medication management skills are most promising and should be considered for implementation in practice.

(J Med Internet Res 2021;23(1):e18901) doi:10.2196/18901

KEYWORDS

eHealth; mHealth; medication adherence; interventions; long-term conditions

Introduction

Long-term medication aims to reduce the risk of disease progression, comorbidity, and mortality [1]. These outcomes will only be reached when patients adhere to their medication. Presumably 50% of all patients with long-term medication are nonadherent, leading to suboptimal treatment outcomes [2,3]. Although improvement on clinical outcomes is the ultimate treatment goal, measuring adherence to long-term medication allows for comparison across chronic conditions.

Medication adherence is defined as the extent to which medication taking behavior corresponds with the medication regimen agreed upon with the health care professional [4]. Medication-taking behavior is influenced by different factors such as experience, beliefs, and culture, making it multifaceted. Moreover, medication-taking behavior can differ between each drug and may change over time. Targeted, timely interventions enhancing medication adherence have therefore become one of the major priorities in health care. Despite efforts, randomized controlled trials have demonstrated limited effectiveness of medication-enhancing interventions [5-9]. Besides, effective interventions differed markedly and did not apply similar intervention strategies, making comparisons or meta-analysis difficult [3,5-7].

eHealth might provide an opportunity to offer accessible, interactive, timely, and feasible medication adherence interventions that require minimal effort from health care providers whose time and resources are limited. eHealth or telemedicine—these words are used interchangeably—is defined as the use of information and communication technology in health care [10]. These technologies can facilitate tailored and interactive solutions such as targeted education, consistent support, and skill acquisition. Thus, the multifaceted and versatile medication-taking behavior can well be targeted by eHealth interventions.

eHealth seems a promising way forward but recent systematic reviews showed conflicting results for eHealth interventions on improving medication adherence [11-14]. These reviews focused on single long-term conditions and have led to fragmented knowledge on the effectiveness and strategies of eHealth interventions. Evidence on eHealth interventions should be clustered to comprehensively investigate effectiveness of eHealth interventions and facilitate generalizability of study findings. Sieben et al [5] and Linn et al [15] found promising results across long-term conditions but the fast developments in eHealth render those results outdated, their definition of eHealth as "internet" was too narrow, and included studies had methodological limitations. Therefore the aim of our systematic review is twofold: (1) to evaluate effectiveness of recent interactive eHealth interventions on medication adherence in adult patients using long-term medication and (2) to describe applied strategies within effective interventions.

Methods

This systematic review adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) statement [16] and was completed according to the registered protocol PROSPERO 2019 CRD42019088873 [17].

Search Strategy and Study Selection

Searches were undertaken in MEDLINE, EMBASE, Cochrane Library, PsycINFO, and Web of Science to identify eligible studies. The search strategy comprised 3 blocks: eHealth, medication adherence, and randomized clinical trial (see Multimedia Appendix 1 for the MEDLINE search). Reference lists and citations of included studies were checked to ensure literature saturation. Titles and abstracts were screened and full-text articles were assessed by 2 researchers (BP and BvdB or BP and JV) independently based on the inclusion criteria below. Discrepancies between researchers were resolved through discussion or by reaching consensus with the third researcher (BvdB or JV).

Eligibility Criteria

Eligible studies fulfilled the following inclusion criteria: (1) randomized controlled trial with a usual care control group; (2) applying an interactive eHealth intervention aimed at the patient or patient's caregiver; (3) medication adherence as primary outcome; (4) a total sample size of at least 50 adult patients using long-term medication as determined by Zwikker et al [18]; and (5) published between 2014 and July 2019. Only peer-reviewed English full-text articles were included. We considered all interventions solely applied over distance as eHealth interventions (eg, online portals, telephone calls). Blended care interventions, where face-to-face contact is combined with online components, were excluded. Interventions were considered interactive if there was tele-feedback regardless by whom on medication adherence (eg, bidirectional text messaging, sending adherence reports). Only validated medication adherence outcomes (ie, objective measures or subjective measures that have been compared to objective measures) were taken into account.

Quality of Evidence

Two researchers (BP and JV) independently assessed the internal validity of included studies using the Cochrane Collaboration's tool for assessing risk of bias [19]. Seven domains were scored as having low (+), high (–), or unclear (?) risk of bias. Because blinding of participants and personnel is hardly feasible in studies evaluating interventions aimed at adherence, this domain was considered high risk (–) for all studies. Studies with a positive score (+) on at least five domains were considered high-quality studies. If relevant information was not reported, the corresponding author was contacted to request additional information. When no additional relevant information was provided, the risk of bias domain was scored as unclear (?).

Data Extraction

A standardized template was made to extract data on study characteristics, eHealth interventions, and medication adherence outcomes. Details of the eHealth interventions were extracted according to the Template for Intervention Description and Replication (TIDieR) checklist [20]. Additionally, the mode of adherence tele-feedback of each eHealth intervention was extracted. We distinguished the following modes of adherence tele-feedback: monitoring device, SMS text messaging, interactive voice response (IVR), mobile app, call, or e-training. Intervention strategies were categorized based on the strategies defined by Lowe et al [21] to support behavior change (ie, strategies focusing on adopting treatment behaviors); to inform and educate; to support (ie, strategies assisting consumers with their medicines use such as peer support); to teach skills; to minimize risk and harms (ie, strategies focusing on preventing or managing adverse events); to facilitate communication or decision making; and to improve health care quality (ie, strategies improving, coordinating, or integrating the delivered care). Only the adherence outcome at study endpoint was extracted where magnitude of the intervention effect would be most apparent. For continuous outcomes Cohen *d* for assessing effect size was calculated if means and SDs were provided [22]. Dichotomous outcomes were recalculated to odds ratio (OR) where OR ≥2.0 is deemed to be a strong OR [23]. Additionally, if the authors reported a statistical significance favoring the intervention group compared to the control group, this was scored positive (+). A negative score (–) means there was no statistically significant difference to report. Data were extracted by one researcher (BP) and checked for accuracy by a second researcher (JV).

Data Analysis

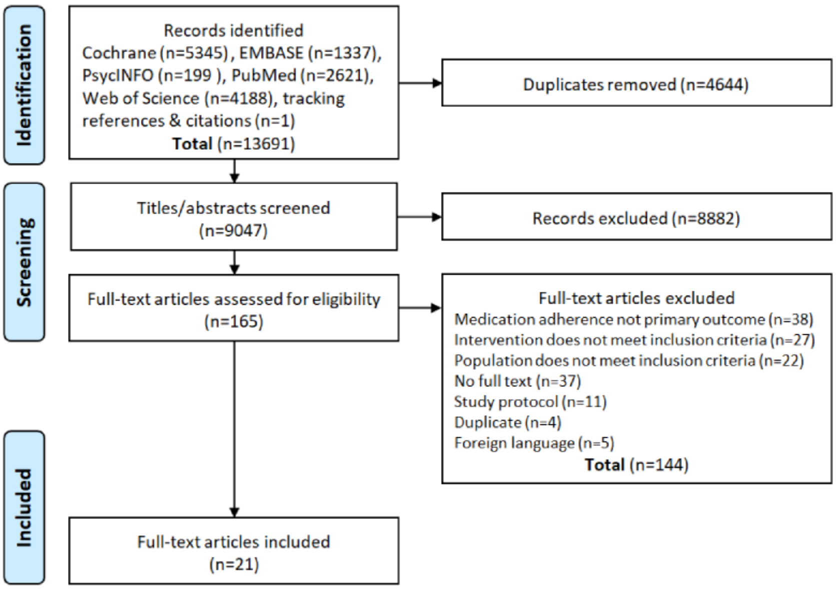
Statistical data pooling was not feasible due to heterogeneity between studies and interventions. Therefore a best evidence synthesis was performed to examine the effectiveness of interactive eHealth interventions on medication adherence. The Cochrane Back Review Group defines 4 levels of evidence: strong, moderate, limited, and conflicting evidence [24]. Strong evidence reflects consistent (ie, 75% or more of the studies report results in the same direction) results among 2 or more high quality studies. Moderate evidence reflects consistent results of 1 high-quality study and 2 or more lower-quality studies. Limited evidence reflects the result of 1 lower-quality study. Conflicting evidence reflects inconsistent results among 2 or more studies. If there were 2 or more high-quality studies, the lower-quality studies were disregarded in the best evidence synthesis. A post-hoc sensitivity analysis was performed to examine the robustness of the best evidence synthesis using a more stringent cut-off score (6 out of 7 instead of 5 out of 7 domains graded as low risk of bias) for determining the quality of the studies.

Results

Search Results

Figure 1 shows a flow diagram of the literature search which provided a total of 9047 publications for screening and yielded 21 articles reporting on 29 interactive eHealth interventions [25-45]. One article, by Reese et al [34], reported on 2 studies. Five studies reported on more than 1 intervention [28,34,35,38,45].

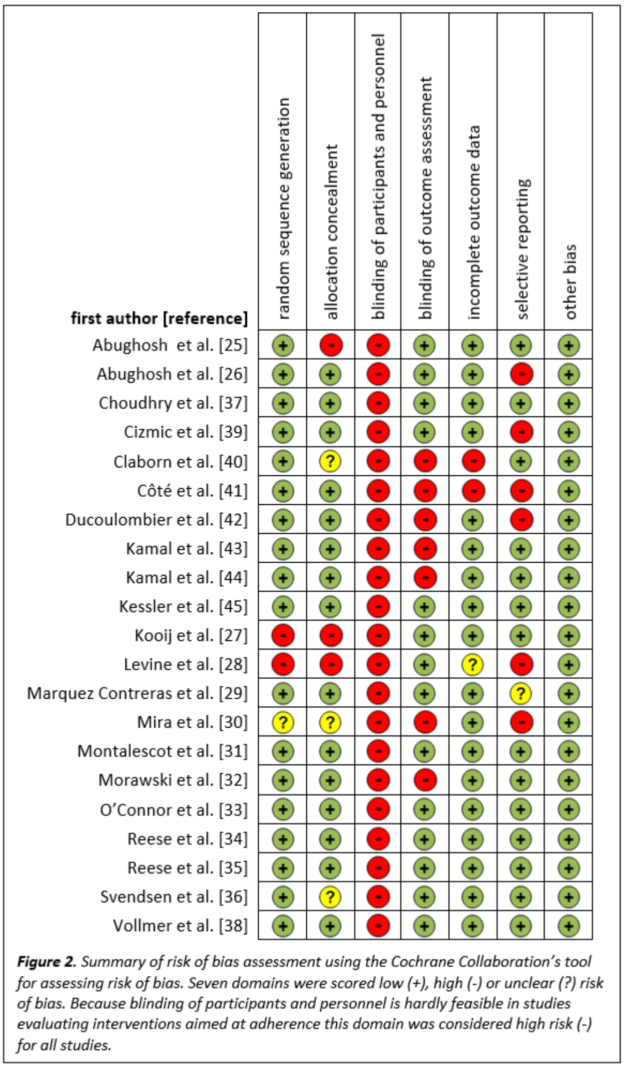
Figure 1. PRISMA flow diagram of study search and selection.



Risk of Bias Assessment

Fifteen studies had a positive score on at least five domains and were regarded high-quality studies as shown in Figure 2. Two studies had the lowest score with 2 out of 7 domains scored as positive.

Figure 2. Summary of risk of bias assessment using the Cochrane Collaboration's tool for assessing risk of bias.



first author [reference]	random sequence generation	allocation concealment	blinding of participants and personnel	blinding of outcome assessment	incomplete outcome data	selective reporting	other bias
Abughosh et al. [25]	+	-	-	+	+	+	+
Abughosh et al. [26]	+	+	-	+	+	-	+
Choudhry et al. [37]	+	+	-	+	+	+	+
Cizmic et al. [39]	+	+	-	+	+	-	+
Claborn et al. [40]	+	?	-	-	-	+	+
Côté et al. [41]	+	+	-	-	-	-	+
Ducoulombier et al. [42]	+	+	-	-	+	-	+
Kamal et al. [43]	+	+	-	-	+	+	+
Kamal et al. [44]	+	+	-	-	+	+	+
Kessler et al. [45]	+	+	-	+	+	+	+
Kooij et al. [27]	-	-	-	+	+	+	+
Levine et al. [28]	-	-	-	+	?	-	+
Marquez Contreras et al. [29]	+	+	-	+	+	?	+
Mira et al. [30]	?	?	-	-	+	-	+
Montalescot et al. [31]	+	+	-	+	+	+	+
Morawski et al. [32]	+	+	-	-	+	+	+
O'Connor et al. [33]	+	+	-	+	+	+	+
Reese et al. [34]	+	+	-	+	+	+	+
Reese et al. [35]	+	+	-	+	+	+	+
Svendsen et al. [36]	+	?	-	+	+	+	+
Vollmer et al. [38]	+	+	-	+	+	+	+

***Figure 2.** Summary of risk of bias assessment using the Cochrane Collaboration's tool for assessing risk of bias. Seven domains were scored low (+), high (-) or unclear (?) risk of bias. Because blinding of participants and personnel is hardly feasible in studies evaluating interventions aimed at adherence this domain was considered high risk (-) for all studies.*

Description of Study Population

Over half of the studies (13/22) included long-term medication for cardiovascular disease, diabetes, or both. Seven studies focused on other, single long-term conditions, leaving 2 studies that looked at any long-term conditions where long-term medication was in use.

The smallest study reported on 70 participants at baseline and the largest study involved 21,752 participants. Because all studies were randomized, baseline characteristics of the different groups were generally the same.

Description of Study Design

Follow-up was short (ie, less than 6 months) in 11 studies and long (at least six months) in 11 studies. The primary medication adherence outcome of each of the studies was mainly assessed objectively using medication monitoring devices, pharmacy prescription data, and serum levels. The remaining 6 studies measured adherence subjectively with validated self-report questionnaires (eg, Immunosuppressant Therapy Adherence Instrument).

Description of eHealth Interventions and Intervention Strategies

Twenty-nine different interactive eHealth interventions were evaluated as shown in Table 1. Most (21/29, 72%) interventions specified using a (mobile) phone for calling, SMS text messaging, or mobile apps.

Most (25/29) interventions were aimed at the patient, 3 interventions were aimed at the caregiver, and another was aimed at either patient or caregiver.

Sixteen interventions were provided through automated software without involvement of a health care professional: 6 mobile apps, 5 monitoring devices, 3 SMS text messages or IVR interventions, and 2 e-training modules through an online portal. Another 7 interventions were provided through automated software in combination with tele-feedback by a health care professional or caregiver: 4 monitoring devices, 2 IVR or SMS text message interventions, 1 e-training. The 6 remaining interventions were telephone calls performed by health care professionals.

Regarding intervention strategies, nearly all (23/29, 79%) interventions aimed at informing and educating patients and just over half (15/29, 52%) sought to support patients by providing assistance and encouragement. All other strategies (eg, teaching skills, facilitating communication or decision making) were less frequently applied (see Multimedia Appendix 2).

Effectiveness of eHealth Interventions on Medication Adherence

Overall, 17 interventions yielded a statistically significant improvement of medication adherence compared to the control group (Table 2). For 14 of these interventions an effect size (Cohen *d*) could also be calculated; 2 interventions reported a large effect size (Cohen $d \geq 0.8$) [25,29], 4 had a medium effect size (Cohen $d \geq 0.5 < 0.8$) [35,43,45], 3 had a small effect size (Cohen $d \geq 0.2 < 0.5$) [26,32,35], and 5 interventions had a negligible effect size (Cohen $d < 0.2$) [27,30,37,38]. For the remaining 3 interventions an OR could be calculated which showed strong odds for becoming adherent in the intervention group (OR ≥ 2.0) [36,39,42].

Table 1. Characteristics of the eHealth interventions.

Study and medication	Adherence inclusion criterion	Intervention arm (n)	Control arm (n)	Follow-up (in weeks)	Mode of adherence telefeedback	Description of the intervention
Levine et al [28]						
Tacrolimus	None	38	50	13	App	Transplant Hero is an interactive alarm to remind patients to take their medications as well as providing educational content.
Tacrolimus	None	20	50	13	App and smart-watch	Transplant Hero (see above) combined with a smartwatch that displayed the reminder notifications.
Cizmic et al [39]						
Bisphosphonates	None	127	118	4	IVR[a]	An IVR call focusing on known reasons for not initiating therapy. If the medication was not picked up 7 days after receiving the call, a reminder letter was sent.
O'Connor et al [33]						
Antihypertensives or medication for lowering blood glucose or cholesterol	None	1220	1158	9	Call	A single-protocol-structured telephone call from an interventionist using positive reinforcement and probing for reasons of nonadherence.
Kessler et al [45]						
Statins	<80%	51	34	26	Device	A wireless pill bottle generated an alert message, sent to the participant, if medication was missed the previous day and at least once in the 2 prior days.
Statins	<80%	46	34	26	Device	A wireless pill bottle generated an automated alert message (see above), sent to the participant and a designated caregiver.
Márquez Contreras et al [29]						
Antihypertensives	None	73	75	52	App	The AlerHTA app aimed to promote health education in hypertension and remind for both appointments and medication intake time.
Montalescot et al [31]						
Apixaban	None	579	583	24	e-Training	An education program consisting of an education booklet, one or more reminder tools chosen by the participant, and access to a telephone clinic.
Reese et al [34]						
Statins	<80%	67	67	13	Device	PROMOTE-1: a wireless pill bottle generated a weekly adherence report in which the patient's adherence was compared to other patients.
Statins	<80%	67	67	13	Device	PROMOTE-2: a wireless pill bottle generated a weekly adherence report.
Statins	<80%	50	50	13	Device	SUPPORT-1: a wireless pill bottle generated a daily adherence report.
Statins	<80%	50	50	13	Device	SUPPORT-2: a wireless pill bottle generated a weekly adherence report.
Statins	<80%	50	50	13	Device	SUPPORT-3: a wireless pill bottle generated an email alert if the patient missed a dose the previous day.
Reese et al [35]						
Tacrolimus	None	40	40	26	Device	A wireless pill bottle generated an alert when medication was due and patients could select additional reminders such as SMS text messages, calls, or emails.

Study and medication	Adherence inclusion criterion	Intervention arm (n)	Control arm (n)	Follow-up (in weeks)	Mode of adherence telefeedback	Description of the intervention
Tacrolimus	None	40	40	26	Device	A wireless pill bottle generated an alert (see above). If adherence decreased to <90% in a 14-day period, the study coordinator would call the patient and notify the involved HCPs[c].
Svendsen et al [36]						
Calcipotriol/betamethasone foam	none	68	66	4	App	An app which provided once-daily reminders and information on number of treatment applications and amount of prescribed foam applied.
Abughosh et al [25]						
RAS[b] inhibitors	<80%	87	99	26	Call	A brief telephone intervention by pharmacists to remind the patients of their overdue refill and to identify potential adherence barriers.
Abughosh et al [26]						
RAS inhibitors	<80%	248	495	26	Call	Six motivational interviewing phone calls by pharmacy students to identify potential adherence barriers and provide guidance to address these barriers.
Choudhry et al [37]						
Antihypertensives or medication for lowering blood glucose or cholesterol	<80%	2038	2040	52	Call	Tailored telephone consultation to develop a shared plan to improve adherence and disease control. At 6 and 9 months progress reports were mailed.
Ducoulombier et al [42]						
Bisphosphonates and strontium ranelate	None	79	85	52	Call	Bimonthly telephone follow-up to motivate patients to maintain good adherence, detect difficulties in compliance, and recall the importance of treatment continuation.
Kooij et al [27]						
Bisphosphonates, RAS inhibitors, and statins	None	2008	2914	52	Call	Telephone counselling 7-21 days after the start of therapy assessing practical and perceptual barriers and providing information and motivation.
Vollmer et al [38]						
RAS inhibitors and statins	<90%	7247	7255	52	IVR	An IVR call when (over)due for a refill providing patient education and refill support.
RAS inhibitors and statins	<90%	7250	7255	52	IVR	In addition to IVR calls (see above), a reminder letter was sent if they were 60-89 days overdue, a call was made if they were ≥90 days overdue, and primary care provider informed. Patients also received a personalized health report, a pill organizer, and bimonthly mailings.
Claborn et al [40]						
Highly active antiretroviral therapy	<95%	47	50	4	e-Training	eLifeSteps: a single-session, self-paced multimedia intervention tackling practical and psychological adherence barriers accompanied with a workbook.
Côté et al [41]						
Immunosuppressants	None	35	35	26	e-Training	Transplant-TAVIE was composed of 3 interactive Web-based sessions by a virtual nurse aimed at developing and reinforcing self-management skills required for medication intake.
Kamal et al [43]						

Study and medication	Adherence inclusion criterion	Intervention arm (n)	Control arm (n)	Follow-up (in weeks)	Mode of adherence telefeedback	Description of the intervention
Preventive medication for stroke	None	100	100	8	SMS text messages	SMS4stroke sent automated customized SMS text message reminders to either patient or caregiver.
Kamal et al [44]						
Statins and antiplatelets	None	99	98	13	IVR and SMS	Daily IVR call services, daily prescription-tailored medication reminders, and once weekly life style modification messages.
Mira et al [30]						
All medication allowed, >2	None	51	49	13	App	A tablet-based medication self-management app (ALICE) with medication reminders and medication information such as pictures, interactions, storage instructions, and common errors in medication use.
Morawski et al [32]						
Antihypertensives	None	209	202	12	App	The MediSafe app is a medication reminder app with additional functions such as adherence reports, tracking of measurements, and peer support.

[a] IVR: interactive voice response.

[b] RAS: renin–angiotensin system.

[c] HCP: health care professional.

Table 2. Adherence measure and medication adherence results of the studies reviewed.

Adherence measure, study, and results on medication adherence	Statistically significant[a]
Serum level (0 - 100)	
Levine et al [28]	
The coefficient of variability (SD/mean × 100) of tacrolimus levels was 33.0 for the intervention group and 32.8 for the control group (Cohen *d* = 0.01).	–
The coefficient of variability was 33.8 for the intervention group and 32.8 for the control group (Cohen *d* = 0.07).	–
Fill first prescription (0 %- 100%)	
Cizmic et al [39]	
49% of the intervention group filled their first prescription compared to 31% of the control group (OR[b] 2.17; 95% CI 1.29-3.67).	+
O'Connor et al [33]	–
84% of the intervention group filled their first prescription compared to 84% of the control group (OR 0.94; 95% CI 0.79-1.11).	–
Bottle openings (0%-100%)	
Kessler et al [45]	
Average daily adherence was 53% for the intervention group and 36% for the control group (Cohen *d* = 0.70).	+
Average daily adherence was 55% for the intervention group and 36% for the control group (Cohen *d* = 0.70).	+
Márquez Contreras et al [29]	
Average daily adherence was 86% for the intervention group and 63% for the control group (Cohen *d* = 4.72).	+
Montalescot et al [31]	
Average daily adherence was 92% for the intervention group and 92% for the control group (Cohen *d* = 0.02).	–
Reese et al [34]	
Average daily adherence was 77% for the intervention group and 75% for the control group.	–
Average daily adherence was 71% for the intervention group and 75% for the control group.	–
Reese et al [34]	
Average daily adherence was 73% for the intervention group and 79% for the control group.	–
Average daily adherence was 75% for the intervention group and 79% for the control group.	–
Average daily adherence was 75% for the intervention group and 79% for the control group.	–
Reese et al [35]	
Average daily adherence (during the final 90 days) was 78% for the intervention group and 55% for the control group (Cohen *d* = 0.37).	+
Average daily adherence (during final 90 days) was 88% for the intervention group and 55% for the control group (Cohen *d* = 0.57).	+
Svendsen et al [36]	
66% of the intervention group was considered adherent compared to 38% of the control group (OR 3.22; 95% CI 1.53-6.80).	+
PDC[c,d] (0% - 100%)	
Abughosh et al [25]	
PDC was 58% for the intervention group and 29% for the control group (Cohen *d* = 1.32).	+
Abughosh et al [26]	
PDC was 66% for the intervention group and 57% for the control group (Cohen *d* = 0.26).	+
Choudhry et al [37]	
PDC was 46% for the intervention group and 42% for the control group (Cohen *d* = 0.12).	+
Ducoulombier et al [42]	+

Adherence measure, study, and results on medication adherence	Statistically significant[a]
65% of the intervention group was considered adherent compared to 33% of the control group (OR 3.71; 95% CI 1.94-7.07).	+
Kooij et al [27]	
PDC was 81% for the intervention group and 76% for the control group (Cohen *d* = 1.34).	+
Vollmer et al [38]	
PDC was 58% for the intervention group and 56% for the control group (Cohen *d* = 2.09).	+
PDC was 59% for the intervention group and 56% for the control group (Cohen *d* = 2.14).	+
AACTGAI[e] (0% - 100%)	
Claborn et al [40]	
Adherence was 81% for the intervention group and 81% for the control group (Cohen *d* = –0.03).	–
ITAS[f] (0-12)	
Côté et al [41]	
Mean ITAS score was 11.7 in the intervention group and 11.3 in the control group (Cohen *d* = 0.30).	–
MMAS[g] (0-8)	
Kamal et al [43]	
Mean MMAS score was 7.4 in the intervention group and 6.7 in the control group (Cohen *d* = 0.62).	+
MMAS (0-8)	
Kamal et al [44]	
Mean MMAS score was 7.3 in the intervention group and 7.1 in the control group (Cohen *d* = 0.03).	–
MMAS-4 (0-8)	
Mira et al [30]	
Mean MMAS score was 7.4 in the intervention group and 7.3 in the control group (Cohen *d* = 0.12; not corrected for baseline).	+
MMAS (0-8)	
Morawski et al [32]	
Mean MMAS score was 6.3 in the intervention group and 5.7 in the control group (Cohen *d* = 0.35).	+

[a] As reported by the authors. + indicates $P<.05$ favoring intervention and – indicates $P>.05$ (no significant difference between groups).

[b] OR: odds ratio.

[c] PDC: percentage of days covered.

[d] All PDC outcomes were based on refill data; pill counts were considered separately.

[e] AACTGAI: Adult AIDS Clinical Trials Group Adherence Instrument.

[f] ITAS: Immunosuppressant Therapy Adherence Instrument.

[g] MMAS: Morisky Medication Adherence Scale.

Further details of the study, population, intervention, and outcomes can be found in the extraction database provided as Multimedia Appendix 5.

The best evidence synthesis (Table 3) showed strong evidence for a positive effect for SMS text messages or IVR, mobile apps, and calls as mode of adherence tele-feedback. The evidence for e-training was weak and for monitoring devices conflicting.

In the post hoc sensitivity analysis the criteria for a high-quality study were more stringent (6 out of 7 domains graded as low risk of bias). The sensitivity analysis showed that the strong evidence for a positive effect for SMS or IVR as mode of adherence tele-feedback remained, whereas the evidence turned to conflicting for interventions delivered through mobile apps and calls (see Multimedia Appendix 3).

The level of evidence of the intervention strategies was also assessed. There was strong evidence for a positive effect of strategies to teach skills, to facilitate communication or decision making, and to improve health care quality. For all other intervention strategies (eg, to support, to inform and educate) there was conflicting evidence (see Multimedia Appendix 4).

Table 3. Results of the best evidence synthesis.

Mode of adherence tele-feedback and quality	Statistically significant[a]	Level of evidence
Monitoring device		Conflicting evidence
9 HQ[b] interventions	+, +, +, +, –, –, –, –, –	
0 LQ[c] interventions		
SMS text messaging or IVR[d]		Strong evidence for a positive effect
5 HQ interventions	+, +, +, +, –	
0 LQ interventions		
Mobile app		Strong evidence for a positive effect
3 HQ interventions	+, +, +	
3 LQ interventions	+, –, –	
Call		Strong evidence for a positive effect
4 HQ interventions	+, +, +, –	
2 LQ interventions	+, +	
e-Training		Moderate evidence for no effect
1 HQ intervention	–	
2 LQ interventions	–, –	

[a] + indicates $P<.05$ favoring intervention; – indicates $P>.05$ (no significant difference between groups). In grading the level of evidence low-quality studies were disregarded when there were 2 or more high-quality studies.

[b] HQ: high quality.

[c] LQ: lower quality.

[d] IVR: interactive voice response.

Discussion

Principal Findings

This systematic review examined the effectiveness of interactive eHealth interventions to improve medication adherence in patients using long-term medication published between 2014 and 2019. A majority, 17/29 interactive interventions, had a statistically significant ($P<.05$) effect on medication adherence. There was strong evidence for a positive effect for interventions using SMS or IVR, mobile apps, and calls as mode of adherence tele-feedback. Intervention strategies to teach skills, to improve health care quality, and to facilitate communication or decision making showed a strong evidence for a positive effect. Overall, this review shows that interactive eHealth interventions can be effective in improving medication adherence.

Interactive eHealth Interventions

This study showed strong evidence for a positive effect on medication adherence of eHealth interventions across various channels, including SMS, IVR, mobile apps, and calls. Our findings add robustness to the positive effect of eHealth interventions provided by previous systematic reviews and meta-analyses [46-51]. Where those authors were cautious with interpreting their findings because of low-quality studies, small sample sizes, and short follow-up, many studies we included were of high quality (22/29), had sample sizes of 100 patients or more (19/29), and follow-up of at least six months (14/29). IVR interventions that included information about health consequences suggest a stronger behavioral change, including medication-taking behavior [51]. This review confirms these findings as the included IVR interventions all contained information on the consequences of (not) taking medication as prescribed. For call, mobile app, and SMS text messaging interventions it remains unclear whether there are intervention elements (eg, content, intervention design, or extent of tailoring) that contribute to increased medication adherence because most eHealth interventions are multicomponent and elements vary widely across interventions [46,48,50].

We found a lack of convincing evidence for interventions applying an electronic monitoring device or e-training. By contrast, van Heuckelum et al [52] found a positive effect for interventions using monitoring device feedback. In our review all interventions coupled their electronic monitoring devices to the same (Way To Health) communication platform which could be a possible explanation. Yet, van Heuckelum et al [52] also included interventions that gave face-to-face feedback on adherence data collected by monitoring devices. They showed that these interventions were effective, whereas those that applied tele-feedback were not. This suggests that feedback on tele-monitoring of medication adherence is best given face-to-face.

Intervention Strategies

To describe intervention strategies among effective interactive eHealth interventions we used Lowe's taxonomy as it is specific for adherence interventions with clear examples for each strategy. Although other taxonomies (eg, Abraham and Michie [53], Demonceau et al [54], Kini and Ho [7]) could have been used, they show many conceptual similarities with Lowe’s taxonomy. Following Lowe’s taxonomy, we provide evidence for interactive adherence interventions aimed at teaching skills such as self-management programs, improving health care quality by coordinating medication adherence care between professionals, and facilitating communication or decision making between patient and health care professional. These results should be interpreted with caution because interventions were multifaceted and combined different strategies. Furthermore, the strategies with the highest level of evidence were also those that were less used. It is not possible to assign success to a single strategy within a multifaceted intervention. Nonetheless, the effective strategies we identified in this review suggest to be good starting points for development or selection of interventions.

Patient Populations

Noteworthy, the included studies in our review using eHealth interventions to address medication adherence reflect 2 distinct patient populations, namely, the large patient population (eg, metabolic and cardiovascular disease) and the population where optimal medication adherence is critical (eg, immunosuppressants, antiretroviral therapy). Applying eHealth to address medication adherence can be advantageous for both populations albeit for different reasons. eHealth interventions can be accessible for large patient populations, giving health care professionals a large outreach with limited resources. For populations where optimal medication adherence is critical, eHealth interventions can be tailored to patients’ specific needs and provide continuous support.

Study Quality

Where others found a lack of high-quality studies and stressed the importance of improving study quality [3,9,15], this review included 15 (out of 23) high-quality studies. The increase in quality presumably is a direct consequence of better reporting and study designs. We chose the Cochrane risk of bias tool (version 1) to assess study quality. This tool focusses on internal validity and does not cover all aspects of study design. We found flaws in study design that were not covered by the Cochrane risk of bias tool such as absence of sample size calculation, selection bias, and disputable (adherence) outcomes. This could have (negatively) affected the implications of the results.

Clinical Implications and Future Research

We clustered evidence of various long-term conditions in our best evidence synthesis to provide a comprehensive overview. This overview is based on the statistically significant effects ($P<.05$) found by the authors and supported by the effect sizes we calculated. Of the 17 statistically significant ($P<.05$) interventions, 9 showed at least a small effect size (Cohen $d \geq 0.2$) and 3 interventions showed strong odds (OR > 2.0) for effect in the intervention group.

The synthesis was limited to medication adherence and did not consider other clinical outcomes. As a result, our findings may not be applicable one-on-one to specific conditions. Next step is to study the identified effective interventions/strategies in specific long-term conditions to ascertain that this may lead to improved medication adherence and other clinical outcomes.

Only postintervention effectiveness on medication adherence was assessed in this review. Whether the found beneficial effects will be maintained over a longer period (>12 months) remains unclear. However, 12/17 effective interventions in our review had a follow-up of at least six months which is considered the shortest period to accurately assess long-term medication adherence [4].

We were surprised to find many interactive eHealth interventions that use technologies published in the 20th century. Although technology changes, applied techniques are very similar. To be able to build upon data and lessons learnt from older technologies, crosslinks between similar techniques need to be made (eg, between SMS text messaging and chat services such as WhatsApp or WeChat).

Technological developments are very fast paced and eHealth interventions continuously change. This high turnaround speed creates a need for study designs that allow continuous evaluation of interventions over a period of at least six months.

In this review intervention exposure ranged from a single call to daily messages for months. To establish a relation between exposure and medication taking behavior change, dose–response studies are called for.

Conclusion

We found that a majority of interactive eHealth interventions are effective in improving adherence to long-term medication. Intervention strategies that improve patient’s treatment involvement and their medication management skills are most promising. While most interactive eHealth interventions were multifaceted, even simple eHealth technologies such as SMS text messaging and telephone calls can be effective in promoting medication adherence in a wide variety of patient populations.

Authors' Contributions

BP developed the PROSPERO protocol which was reviewed by JV, AL, HvO, MV, SvD, and BvdB. BP developed and conducted search strategy and screening and inclusion of the studies, designed extraction template, extracted data, and drafted the manuscript. Rating study quality and revision of subsequent drafts of the manuscript were done by BP and JV. JV and BvdB also performed article screening. CB, AL, HvO, MV, SvD, and BvdB critically reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[DOC File , 30 KB - jmir_v23i1e18901_app1.doc]

Multimedia Appendix 2
Overview of the intervention strategies present in each e-Health intervention.
[PNG File , 27 KB - jmir_v23i1e18901_app2.png]

Multimedia Appendix 3
Sensitivity analysis.
[DOC File , 33 KB - jmir_v23i1e18901_app3.doc]

Multimedia Appendix 4
Level of evidence of intervention strategies.
[DOC File , 37 KB - jmir_v23i1e18901_app4.doc]

Multimedia Appendix 5
Extraction database.
[XLSX File (Microsoft Excel File), 121 KB - jmir_v23i1e18901_app5.xlsx]

References

1. Christensen A. Patient Adherence to Medical Treatment Regimens. New Haven, CT: Yale University Press; 2004.
2. Corrao G, Parodi A, Nicotra F, Zambon A, Merlino L, Cesana G, et al. Better compliance to antihypertensive medications reduces cardiovascular risk. J Hypertens 2011 Mar;29(3):610-618. [doi: 10.1097/HJH.0b013e328342ca97] [Medline: 21157368]
3. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;11:CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
4. Przemyslaw K, Andrzejczyk A, Clyne W, De GS, Demonceau J, Dobbels F. Ascertaining Barriers for Compliance: policies for safe, effective and cost-effective use of medicines in Europe - Final Report of The ABC Project. 2012. URL: http://abcproject.eu/img/ABC%20Final.pdf [accessed 2020-12-16]
5. Sieben A, Bredie SJHB, van Laarhoven CJHMK, Schoonhoven L, Burger DM, van Onzenoort HAW. [e-Health interventions and improvement in treatment adherence]. Ned Tijdschr Geneeskd 2014;158:A8460. [Medline: 25534266]
6. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RCM, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 4;157(11):785-795. [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
7. Kini V, Ho PM. Interventions to Improve Medication Adherence: A Review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
8. Wilhelmsen NC, Eriksson T. Medication adherence interventions and outcomes: an overview of systematic reviews. Eur J Hosp Pharm 2019 Jul;26(4):187-192. [doi: 10.1136/ejhpharm-2018-001725] [Medline: 31338165]
9. Conn VS, Ruppar TM. Medication adherence outcomes of 771 intervention trials: Systematic review and meta-analysis. Prev Med 2017 Jun;99:269-276. [doi: 10.1016/j.ypmed.2017.03.008] [Medline: 28315760]
10. Ryu S. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). Healthc Inform Res 2012;18(2):153. [doi: 10.4258/hir.2012.18.2.153]
11. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
12. Wang Z, Zhu Y, Cui L, Qu B. Electronic Health Interventions to Improve Adherence to Antiretroviral Therapy in People Living With HIV: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14404 [FREE Full text] [doi: 10.2196/14404] [Medline: 31621641]
13. Stevenson JK, Campbell ZC, Webster AC, Chow CK, Tong A, Craig JC, et al. eHealth interventions for people with chronic kidney disease. Cochrane Database Syst Rev 2019 Aug 06;8:CD012379. [doi: 10.1002/14651858.CD012379.pub2] [Medline: 31425608]
14. Jeminiwa R, Hohmann L, Qian J, Garza K, Hansen R, Fox BI. Impact of eHealth on medication adherence among patients with asthma: A systematic review and meta-analysis. Respir Med 2019 Mar;149:59-68. [doi: 10.1016/j.rmed.2019.02.011] [Medline: 30803887]
15. Linn AJ, Vervloet M, van DL, Smit EG, Van WJCM. Effects of eHealth interventions on medication adherence: a systematic review of the literature. J Med Internet Res 2011;13(4):e103 [FREE Full text] [doi: 10.2196/jmir.1738] [Medline: 22138112]

16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med 2009;3(3):e123-e130 [FREE Full text] [Medline: 21603045]
17. Pouls B, Van den Bemt BJF, Vriezekolk J. Effectiveness of eHealth interventions on medication adherence in adults with chronic medication: a systematic review. 2019 Jun 19. URL: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=88873 [accessed 2020-12-16]
18. Zwikker HE, van den Bemt BJ, Vriezekolk JE, van den Ende CH, van Dulmen S. Psychosocial predictors of non-adherence to chronic medication: systematic review of longitudinal studies. Patient Prefer Adherence 2014;8:519-563 [FREE Full text] [doi: 10.2147/PPA.S47290] [Medline: 24851043]
19. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
20. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687 [FREE Full text] [Medline: 24609605]
21. Lowe D, Ryan R, Santesso N, Hill S. Development of a taxonomy of interventions to organise the evidence on consumers' medicines use. Patient Educ Couns 2011 Nov;85(2):e101-e107. [doi: 10.1016/j.pec.2010.09.024] [Medline: 21036505]
22. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
23. van der Windt DA, Thomas E, Pope DP, de Winter AF, Macfarlane GJ, Bouter LM, et al. Occupational risk factors for shoulder pain: a systematic review. Occup Environ Med 2000 Jul;57(7):433-442 [FREE Full text] [doi: 10.1136/oem.57.7.433] [Medline: 10854494]
24. Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 Updated Method Guidelines for Systematic Reviews in the Cochrane Back Review Group. Spine 2009;34(18):1929-1941. [doi: 10.1097/brs.0b013e3181b1c99f]
25. Abughosh SM, Wang X, Serna O, Henges C, Masilamani S, Essien EJ, et al. A Pharmacist Telephone Intervention to Identify Adherence Barriers and Improve Adherence Among Nonadherent Patients with Comorbid Hypertension and Diabetes in a Medicare Advantage Plan. J Manag Care Spec Pharm 2016 Jan;22(1):63-73 [FREE Full text] [doi: 10.18553/jmcp.2016.22.1.63] [Medline: 27015053]
26. Abughosh S, Wang X, Serna O, Esse T, Mann A, Masilamani S, et al. A Motivational Interviewing Intervention by Pharmacy Students to Improve Medication Adherence. J Manag Care Spec Pharm 2017 May;23(5):549-560 [FREE Full text] [doi: 10.18553/jmcp.2017.23.5.549] [Medline: 28448784]
27. Kooij MJ, Heerdink ER, van Dijk L, van Geffen ECG, Belitser SV, Bouvy ML. Effects of Telephone Counseling Intervention by Pharmacists (TelCIP) on Medication Adherence; Results of a Cluster Randomized Trial. Front Pharmacol 2016;7:269 [FREE Full text] [doi: 10.3389/fphar.2016.00269] [Medline: 27625605]
28. Levine D, Torabi J, Choinski K, Rocca JP, Graham JA. Transplant surgery enters a new era: Increasing immunosuppressive medication adherence through mobile apps and smart watches. Am J Surg 2019 Jul;218(1):18-20. [doi: 10.1016/j.amjsurg.2019.02.018] [Medline: 30799019]
29. Márquez Contreras E, Márquez Rivero S, Rodríguez García E, López-García-Ramos L, Pastoriza Vilas JC, Baldonedo Suárez A, Compliance Group of Spanish Society of Hypertension (SEH-LELHA). Specific hypertension smartphone app to improve medication adherence in hypertension: a cluster-randomized trial. Curr Med Res Opin 2018 Nov 15;35(1):1-15. [doi: 10.1080/03007995.2018.1549026] [Medline: 30431384]
30. Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014;16(4):e99 [FREE Full text] [doi: 10.2196/jmir.3269] [Medline: 24705022]
31. Montalescot G, Brotons C, Cosyns B, Crijns HJ, D'Angelo A, Drouet L, AEGEAN Study Investigators. Educational Impact on Apixaban Adherence in Atrial Fibrillation (the AEGEAN STUDY): A Randomized Clinical Trial. Am J Cardiovasc Drugs 2020 Feb;20(1):61-71 [FREE Full text] [doi: 10.1007/s40256-019-00356-2] [Medline: 31243691]
32. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a Smartphone Application With Medication Adherence and Blood Pressure Control: The MedISAFE-BP Randomized Clinical Trial. JAMA Intern Med 2018 Jun 01;178(6):802-809. [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
33. O'Connor PJ, Schmittdiel JA, Pathak RD, Harris RI, Newton KM, Ohnsorg KA, et al. Randomized trial of telephone outreach to improve medication adherence and metabolic control in adults with diabetes. Diabetes Care 2014 Dec;37(12):3317-3324 [FREE Full text] [doi: 10.2337/dc14-0596] [Medline: 25315207]
34. Reese PP, Kessler JB, Doshi JA, Friedman J, Mussell AS, Carney C, et al. Two Randomized Controlled Pilot Trials of Social Forces to Improve Statin Adherence among Patients with Diabetes. J Gen Intern Med 2016 Apr;31(4):402-410 [FREE Full text] [doi: 10.1007/s11606-015-3540-y] [Medline: 26585957]
35. Reese PP, Bloom RD, Trofe-Clark J, Mussell A, Leidy D, Levsky S, et al. Automated Reminders and Physician Notification to Promote Immunosuppression Adherence Among Kidney Transplant Recipients: A Randomized Trial. Am J Kidney Dis 2017 Mar;69(3):400-409. [doi: 10.1053/j.ajkd.2016.10.017] [Medline: 27940063]
36. Svendsen MT, Andersen F, Andersen KH, Pottegård A, Johannessen H, Möller S, et al. A smartphone application supporting patients with psoriasis improves adherence to topical treatment: a randomized controlled trial. Br J Dermatol 2018 Nov;179(5):1062-1071. [doi: 10.1111/bjd.16667] [Medline: 29654699]

37. Choudhry NK, Isaac T, Lauffenburger JC, Gopalakrishnan C, Lee M, Vachon A, et al. Effect of a Remotely Delivered Tailored Multicomponent Approach to Enhance Medication Taking for Patients With Hyperlipidemia, Hypertension, and Diabetes: The STIC2IT Cluster Randomized Clinical Trial. JAMA Intern Med 2018 Sep 01;178(9):1182-1189 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3189] [Medline: 30083727]
38. Vollmer WM, Owen-Smith AA, Tom JO, Laws R, Ditmer DG, Smith DH, et al. Improving adherence to cardiovascular disease medications with information technology. Am J Manag Care 2014 Nov;20(11 Spec No. 17):SP502-SP510 [FREE Full text] [Medline: 25811824]
39. Cizmic AD, Heilmann RMF, Milchak JL, Riggs CS, Billups SJ. Impact of interactive voice response technology on primary adherence to bisphosphonate therapy: a randomized controlled trial. Osteoporos Int 2015 Aug;26(8):2131-2136. [doi: 10.1007/s00198-015-3116-z] [Medline: 25956282]
40. Claborn KR, Leffingwell TR, Miller MB, Meier E, Stephens JR. Pilot study examining the efficacy of an electronic intervention to promote HIV medication adherence. AIDS Care 2014;26(3):404-409. [doi: 10.1080/09540121.2013.824534] [Medline: 23909858]
41. Côté J, Fortin M, Auger P, Rouleau G, Dubois S, Boudreau N, et al. Web-Based Tailored Intervention to Support Optimal Medication Adherence Among Kidney Transplant Recipients: Pilot Parallel-Group Randomized Controlled Trial. JMIR Form Res 2018 Jul 19;2(2):e14 [FREE Full text] [doi: 10.2196/formative.9707] [Medline: 30684400]
42. Ducoulombier V, Luraschi H, Forzy G, Vandecandelaere M, Houvenagel E. Contribution of phone follow-up to improved adherence to oral osteoporosis treatment. Am J Pharm Benefits 2015;7(3):e81-e89.
43. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
44. Kamal AK, Khalid W, Muqeet A, Jamil A, Farhat K, Gillani SRA, et al. Making prescriptions. PLoS One 2018;13(12):e0197671 [FREE Full text] [doi: 10.1371/journal.pone.0197671] [Medline: 30571697]
45. Kessler JB, Troxel AB, Asch DA, Mehta SJ, Marcus N, Lim R, et al. Partners and Alerts in Medication Adherence: A Randomized Clinical Trial. J Gen Intern Med 2018 Sep;33(9):1536-1542 [FREE Full text] [doi: 10.1007/s11606-018-4389-7] [Medline: 29546659]
46. DeKoekkoek T, Given B, Given CW, Ridenour K, Schueller M, Spoelstra SL. mHealth SMS text messaging interventions and to promote medication adherence: an integrative review. J Clin Nurs 2015 Oct;24(19-20):2722-2735. [doi: 10.1111/jocn.12918] [Medline: 26216256]
47. Ng R, Carter SR, El-Den S. The impact of mobile applications on medication adherence: a systematic review. Transl Behav Med 2019 Aug 06:ibz125. [doi: 10.1093/tbm/ibz125] [Medline: 31384950]
48. Palacio A, Garay D, Langer B, Taylor J, Wood BA, Tamariz L. Motivational Interviewing Improves Medication Adherence: a Systematic Review and Meta-analysis. J Gen Intern Med 2016 Aug 9;31(8):929-940 [FREE Full text] [doi: 10.1007/s11606-016-3685-3] [Medline: 27160414]
49. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile Apps for Increasing Treatment Adherence: Systematic Review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
50. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile Telephone Text Messaging for Medication Adherence in Chronic Disease: A Meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
51. Tsoli S, Sutton S, Kassavou A. Interactive voice response interventions targeting behaviour change: a systematic literature review with meta-analysis and meta-regression. BMJ Open 2018 Feb 24;8(2):e018974. [doi: 10.1136/bmjopen-2017-018974] [Medline: 29478016]
52. van Heuckelum M, van den Ende CHM, Houterman AEJ, Heemskerk CPM, van Dulmen S, van den Bemt BJF. The effect of electronic monitoring feedback on medication adherence and clinical outcomes: A systematic review. PLoS One 2017;12(10):e0185453 [FREE Full text] [doi: 10.1371/journal.pone.0185453] [Medline: 28991903]
53. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
54. Demonceau J, Ruppar T, Kristanto P, Hughes DA, Fargher E, Kardas P, et al. Identification and assessment of adherence-enhancing interventions in studies assessing medication adherence through electronically compiled drug dosing histories: a systematic literature review and meta-analysis. Drugs 2013 May;73(6):545-562. [doi: 10.1007/s40265-013-0041-3] [Medline: 23588595]

Abbreviations

AACTG: Adult AIDS Clinical Trials Group Adherence Instrument
HQ: high quality
ITAS: Immunosuppressant Therapy Adherence Instrument
IVR: interactive voice response

LQ: lower quality
MMAS: Morisky Medication Adherence Scale
OR: odds ratio
PDC: percentage of days covered
RAS: renin–angiotensin system
TIDieR: Template for Intervention Description and Replication

Edited by G Eysenbach; submitted 27.03.20; peer-reviewed by L Carmona, R Ramos, M Fleming, T Patel, JK Low, M Taitel; comments to author 12.06.20; revised version received 06.08.20; accepted 30.10.20; published 08.01.21.

Please cite as:
Pouls BPH, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HAW, Vervloet M, van Dulmen S, van den Bemt BJF
Effect of Interactive eHealth Interventions on Improving Medication Adherence in Adults With Long-Term Medication: Systematic Review
J Med Internet Res 2021;23(1):e18901
URL: https://www.jmir.org/2021/1/e18901
doi:10.2196/18901
PMID:33416501

Review

Smartphone-Based Interventions and Internalizing Disorders in Youth: Systematic Review and Meta-analysis

Adrian Buttazzoni[1], MSc; Keshbir Brar[2], BSc; Leia Minaker[1], PhD

[1]School of Planning, University of Waterloo, Waterloo, ON, Canada

[2]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Adrian Buttazzoni, MSc
School of Planning
University of Waterloo
200 University Avenue W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 46564
Email: anbuttazzoni@uwaterloo.ca

Abstract

Background: Mental health disorders in youth are a global issue that have important implications for the future quality of life and morbidity of affected individuals. In the context of public health initiatives, smartphone-based interventions have been suggested to hold the potential to be an effective strategy to reduce the symptoms of mental health disorders in youth; however, further evaluation is needed to confirm their effectiveness. This systematic review and meta-analysis documents and synthesizes existing research on smartphone-based interventions targeting internalizing disorders in youth populations.

Objective: This study aims to synthesize existing research on smartphone-based interventions targeting internalizing disorders in youth populations.

Methods: PubMed and SCOPUS were searched in 2019, and 4334 potentially relevant articles were found. A total of 12 studies were included in the final synthesis. We used the Hedges *g* meta-analysis approach and a random effects model for analysis.

Results: The results of this review note that depression and anxiety are the most commonly targeted symptoms, and unlike other similar topics, most studies reviewed were linked to a proven treatment. The overall pooled effect from the meta-analysis showed small but significant effects (κ=12; N=1370; Hedges g=0.20; 95% CI 0.02-0.38) for interventions in reducing the symptoms of internalizing disorders. In total, 4 subgroup analyses examining specific symptoms and intervention styles found varied small significant and nonsignificant effects.

Conclusions: Future research should focus on developing robust evaluative frameworks and examining interventions among more diverse populations and settings. More robust research is needed before smartphone-based interventions are scaled up and used at the population level to address youth internalizing disorders.

(J Med Internet Res 2021;23(1):e16490) doi:10.2196/16490

KEYWORDS

mental health; meta-analysis; mobile phone; smartphone; systematic review; youth

Introduction

Globally, between 10% and 20% of children and adolescents experience mental illness, with approximately half of all mental illnesses known to begin by the age of 14 years [1]. Poor mental health during these developmental years has been linked with unfavorable outcomes regarding employment, relationships, family formation, and health and disability in early adulthood [2]. Childhood and adolescence are key life stages for interventions with regard to mental health and well-being. However, at present, the screening and support for youth mental health issues in many primary health care systems continue to be inadequate, as even optimistic estimates suggest that only 50% of illnesses are detected by doctors [3]. New and innovative strategies are needed to help address mental health issues in youth. Smartphone-based interventions have been increasingly employed to track symptoms and provide support for individuals on a range of related health issues, such as concussions [4].

Taking this potential and the relatively high digital and tech literacy of younger populations into consideration, there has been a recent growth in the interest of delivering mental health interventions to youth cohorts via smartphones [5].

Internalizing disorders are those in which individuals tend to express distress inwardly, such as anxiety disorders and mood disorders (eg, major depressive disorder [6]). Anxiety is one of the most common disorders in young people and is estimated to affect 4% to 20% of children and adolescents [7], whereas depression is said to affect 2% to 8% of children and adolescents [8] and is a leading cause of disability globally [9]. Similarly, stress is often internalized and can significantly impact an individual's coping abilities, self-esteem, and social relations [10], whereas insomnia (ie, sleep problems) has a robust relationship with depression [11] and other internalizing symptoms including anxiety [12]. Disorders can begin early in childhood and can develop into chronic conditions that negatively impact an individual's relationships, development, and daily functioning in the near future [13]. These disorders are associated with functional impairment, increased risk of depression and suicide [14], and substance abuse issues [15] in the long term. As a result, internalizing disorders carry the potential for high societal burdens [16].

Recently, smartphones have become an essential tool in the targeted support, management, and monitoring of mental health disorders. Common mental health intervention strategies using smartphones include text messaging services [17] and smartphone apps [18]. Growing evidence supports the feasibility and potential of smartphone-based interventions to address mental health issues and disorders [19]. For example, positive effects have been observed regarding memory training for older adults [20] and for people with attention-deficit or hyperactivity disorder [21]. Consequently, smartphones are being increasingly used to address mental health issues of youth and adolescent populations [22]. However, with an increase in the use of smartphones as an intervention delivery strategy, there is a need to improve the evaluation aspect of such interventions [23].

Justification for Review and Meta-Analysis

A recent systematic review of meta-analyses that focused on internet- and mobile phone–based interventions for mental and somatic conditions among children found 8 relevant meta-analyses [24]. Of the included papers, 5 focused on web-based or computerized interventions, one was primarily concerned with psychological-based interventions and one searched for but did not include phone-based interventions. The single meta-analysis that included smartphone-based interventions only analyzed one mobile phone–based intervention study [25]. Another recent meta-review [22] broadly cataloged and synthesized existing reviews of all types focusing on digital health interventions for young people with mental health problems. This review did not report any other meta-analyses (although 2 additional relevant scoping reviews and 1 additional systematic review were reported) beyond those documented by Domhardt et al [24]. To date, no meta-analysis has been conducted that has sought to quantitatively, and solely, evaluate the effectiveness of smartphone apps with regard to mental health (specifically, internalizing disorders) of youth populations. This review and meta-analysis are the first to aim to address these points and present an evaluation.

Review Question and Objective

The primary research question guiding this review and meta-analysis is as follows: What are the study designs, intervention features (review), and effectiveness (meta-analysis) of smartphone-based interventions that aim to minimize or reduce the symptoms of youth internalizing disorders? By applying the population, intervention, comparison, outcome, and context model of Petticrew and Roberts [26], we operationalized our meta-analysis research question as presented in Textbox 1.

Textbox 1. Population, intervention, comparison, outcome, and context review research question breakdown (criteria and description).

- Population: youth (in general, <18 years old; however, in some articles, youth was defined as ≤24 years)
- Intervention: smartphone-based interventions specifically targeting an internalizing disorder in a youth population
- Comparison: control versus intervention, group 1 versus group 2, time 1 versus time 2, etc
- Outcome of interest: effectiveness of intervention in reducing symptoms or intensity of internalizing disorder (ie, anxiety, depression, insomnia, stress)
- Context: any

Following this question, the principal objective of the meta-analysis is to quantitatively evaluate the effectiveness of smartphone-based interventions with regard to youth internalizing disorders.

Methods

Search Strategy

Searches of electronic databases were used to identify and document the articles presented in this review and the meta-analysis in June 2019. Different variations of, and other common terms used for, each focal concept (ie, smartphone, internalizing disorder, and youth population) were discussed and developed by the authors, whereas specific terms were truncated as necessary (Textbox 2).

Textbox 2. Search strategy outline.

- Smartphone

“cell phone” OR “cellular phone” OR “mHealth” OR “mobile health” OR “mobile phone” OR “phone” OR “SMS” OR “short message service” OR “smartphone” OR “text” AND

- Internalizing disorder

“anxiety” OR “depression” OR “internalizing disorder*” OR “internalising disorder*” OR “insomnia” OR “stress” AND

- Youth population

“adolescent*” OR “child*” OR “teen*” OR “youth” OR “young adult”

When deciding on which electronic databases to select for this review, we needed to ensure that content from the fields of behavioral science, pediatrics, psychology, and public health are captured. Therefore, to best incorporate this diversity in content and research areas, we conducted the search strategy in 2 interdisciplinary databases—PubMed and SCOPUS.

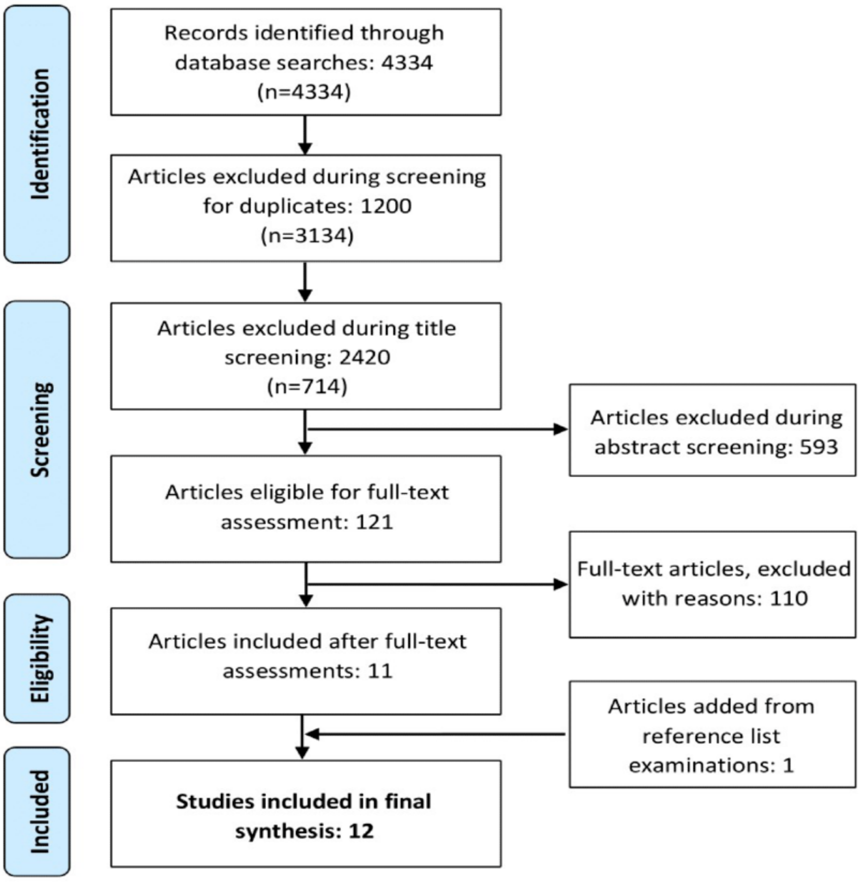
Eligibility Criteria

A total of 5 specific criteria were applied to the article search. The criteria stipulated that each article must (1) be focused primarily on the implementation and evaluation of a smartphone-based intervention (ie, not web-based strategies, social media focused initiatives, etc); (2) have reported some description of the study design and sample, as well as the intervention strategy, implementation, evaluation, and targeted outcome of the intervention strategy; (3) have been a primary research study (ie, not a review, feasibility or acceptability study, proposal, technical report etc) in the design of a randomized controlled trial (RCT), case control, cohort, or cohort analytic design; (4) have focused on and reported the outcome for one or more internalizing disorder outcomes (ie, anxiety, depression, insomnia, stress); and (5) be written in English. There were no geographic or publishing date restrictions placed on the search.

Study Selection and Review Process

Initial searches of PubMed (n=1726) and SCOPUS (n=2608) returned a total of 4334 results (Figure 1). Removal of duplicates resulted in 1200 titles being discarded. The remaining 3134 titles were vetted next, after which 2420 articles were excluded. Abstracts were then screened, resulting in 593 more records being removed. After full-text assessments of the final 121 articles were completed, a total of 11 papers were deemed to meet all of the eligibility criteria. Having included studies that primarily examined adolescents but were mixed with young adult populations, we decided to include 3 studies [27-29] with similar sample demographics but had a mean sample age >18 years and did not explicitly refer to their participants as *adolescents* or *youth*. The most common reasons for exclusion in the full-text assessment phase were feasibility studies lacking a formal assessment, proposal articles only presenting concepts, research papers not including a youth population, and primary studies not reporting an internalizing disorder outcome. Of the total, 1 paper [30] met all of the above conditions; however, its reporting of the outcome of interest was not precise enough to be included in the analysis. Reference list searches of the included articles added 1 additional article to the review. Eventually, 12 studies were documented and synthesized in the systematic review and subsequently included in the meta-analysis. Initial title and abstract vetting was conducted by a team member (AB) with a second member (LM) spot checking the abstract vetting to ensure the consistency of the process. Similar protocols were carried out for the later full-text assessment and data extraction phases (AB and LM).

Figure 1. Flow diagram of study selection processes.



Systematic Review Data Extraction

Systematic review data extractions and meta-analysis coding data for all the included articles are available in the Multimedia Appendix 1. The following information was gathered from each study: citation information, study design information (eg, primary outcome evaluated), smartphone-based intervention details (eg, supporting resources, delivery methods, length), sample details, and internalizing disorder variable information (eg, outcome measure used).

Variables and Concepts Included in Meta-Analysis

Internalizing disorders are often defined as any depressive and anxiety condition [31]. However, other definitions of internalizing disorders have expanded the concept to include panic and stress [32,33], whereas insomnia has become an important symptom of depressive and anxiety disorders [31]. Therefore, this meta-analysis conceptualized internalizing disorders as *anxiety*, *depression*, *insomnia*, or *stress*. *Smartphone-based interventions* were defined as any app, initiative, program, promotion, service, or tool that is based or run primarily through a smartphone.

Meta-Analysis Coding

In addition to the data extractions for the systematic review portion of this manuscript, identified studies pooled in the meta-analysis were also coded to include statistical data. On the basis of the availability of reported results from each study, statistics coded specifically for the meta-analysis included intervention effect sizes (eg, available between- or within-group *d* values and their corresponding SDs), results of *t* tests (eg, differences in means) in cases where no *d* values were reported, scale reliability scores, scale anchors, data source report (eg, self, other), and, if applicable, any information pertaining to multiple measures being employed in the same study.

Meta-Analysis Approach and Statistical Procedures

To guide the meta-analysis, we drew from the outline for effect size meta-analyses of Hunter and Schmidt [34] and chose to use the Hedges g approach. In traditional *d* value meta-analysis approaches, such as the *bare bones* method, no corrections are made for any artifacts other than sampling error [34]; however, given some of the small sample sizes of the included studies, we elected to follow previous recommendations [35] which stipulate using Hedges g as a correction for effect size—a method that pools weighted SDs unlike Cohen *d*. Similar to others [36], we elected to use a more conservative approach given the relatively small number of primary articles being used to estimate the pooled intervention effects. The reported or calculated effect size for the interventions was used to assess the between-group or within-group effectiveness of the interventions on reducing or minimizing internalizing disorder symptoms. Importantly, among the 8 between-group studies, 5 were randomized. Once the *d* values from each study (*uncorrected effect sizes*) were derived, we applied the Hedges g formula to calculate *corrected effect sizes* [35]. A random effects model was used for this meta-analysis because, unlike fixed effects models, it allows for potential variation in underlying parameters. As such, random effects models have been argued to more accurately reflect the real-world heterogeneity of effects [37].

As a part of our conservative approach, all intervention follow-up values used in the meta-analysis were the last date (eg, 2-week and 4-week follow-ups; 4-week selected) reported in each study. Statistics calculated in the meta-analysis included the mean unweighted observed effect, mean effect size weighted by sample size, SD of the sample-weighted effect, corrected weighted variance of the effect (ie, Hedges g), and percentage of variance because of sampling error. We also calculated the same statistics for 4 separate subgroups (specific symptoms and intervention style—2 subgroups each) of the studies. In cases where multiple outcomes of interest were reported in a single study, we calculated the average standardized difference across variables to ascertain the overall effect size [34]. If no effect size data were provided, we followed the method by Morris and DeShon [38] of pooling pre- and posttest means and dividing them by the pretest SD to calculate effects. Sensitivity analyses following the *Cochrane Handbook for Systematic Reviews of Interventions* method [39] were used to calculate effect sizes for articles that only reported baseline and follow-up means (ie, no changes in means data) and to subsequently impute an SD of the change for the experimental group. All analyses were completed in Microsoft Excel.

Quality and Risk of Bias Assessments

Quality assessments (QAs) of the 12 studies reviewed and analyzed are shown in Table 1. To conduct the assessments, we used the Effective Public Health Practice Project (EPHPP) QA tool [40]. Global ratings for each study were developed following the EPHPP dictionary guidelines; 2 reviewers (AB and KB) first calculated independent scores for each article and subsequently evaluations were compared [41]. Comparisons of the evaluations were undertaken to address and resolve any grading variability and other interpretation differences between the 2 reviewers. Reviewers graded each article on its selection bias, study design, confounders, blinding, data collection, withdrawals, and dropouts. Ex-post comparisons of scores between the 2 separately completed individual assessments resulted in an interrater reliability for the QAs of over 90%. The final results of the QA examination were mixed with 4 articles possessing a strong global rating, 5 with a moderate rating, and 3 with a weak rating.

Table 1. Quality assessment of the articles included in meta-analysis.

Study (reference)	Selection bias	Study design	Confounders	Blinding	Data collection	Withdrawals and dropouts	Global rating[a]
Clarke et al 2016 [27]	Moderate	Moderate	Moderate	Weak	Strong	Weak	Weak
Grassi et al 2009 [28]	Moderate	Moderate	Strong	Weak	Strong	Weak	Weak
Kauer et al 2012 [42]	Moderate	Strong	Strong	Moderate	Strong	Strong	Strong
Lee et al 2013 [43]	Weak	Strong	Strong	Moderate	Strong	Weak	Weak
Ranney et al 2016 [44]	Moderate	Moderate	Strong	Weak	Strong	Strong	Moderate
Ranney et al 2018 [45]	Moderate	Strong	Strong	Moderate	Strong	Strong	Strong
Reid et al 2011 [46]	Moderate	Strong	Strong	Moderate	Strong	Moderate	Strong
Stallard et al 2018 [47]	Moderate	Moderate	Strong	Weak	Strong	Strong	Moderate
Takahashi et al 2019 [29]	Moderate	Moderate	Strong	Weak	Strong	Strong	Moderate
Werner-Seidler et al 2019 [48]	Strong	Moderate	Strong	Weak	Strong	Moderate	Moderate
Whittaker et al 2017 [49]	Moderate	Strong	Strong	Strong	Strong	Strong	Strong
Worthen-Chaudhari et al 2017 [50]	Moderate	Moderate	Moderate	Weak	Strong	Strong	Moderate

[a] Per the global rating system, strong: no weak ratings; moderate: 1 weak rating; and weak: 2 or more weak ratings.

More comprehensive risk of bias assessments was also completed for each study included in this review and meta-analysis. A total of 2 separate tools were employed in this regard: nonrandomized intervention papers' risk of bias was assessed using the ROBINS-I (Risk Of Bias In Non-randomised Studies-of Interventions) tool for nonrandomized studies of interventions [51], whereas randomized interventions' (ie, RCTs) risk of bias was assessed using the Cochrane Collaboration tool for assessing risk of bias in randomized trials [52]. Risk of bias assessments of the nonrandomized interventions concluded with all 7 papers being scored as at a *serious risk of bias* overall, whereas evaluations of the randomized trials resulted in 4 of 5 papers being rated as at a *low risk of bias*. Regarding the former result, this trend was largely a consequence of the rating system stipulating that one rating of *serious risk* in any category will be scored with an overall *serious risk of bias* judgment for an article [51]. More specifically, the most common source of potential bias among nonrandomized studies was related to uncertainties regarding the blinding of assessors with regard to the allocation or intervention status of participants, which is not possible in single group interventions (4 out of 7 papers in this portion of the review). Visual illustrations of the results can be found in Multimedia Appendices 2 and 3.

Results

General Characteristics of Included Studies

The characteristics of the 12 included studies are presented in Multimedia Appendix 1. In total, 5 studies were conducted in Australia or New Zealand, 3 in the United States, 2 in Europe, and 2 in Asia. Regarding the evaluated sample populations, 10 of the 12 studies examined between 20 and 120 participants (mean 114), 8 studies had a majority of female participants, 2 had an even split, and 2 had a majority of male participants. Depressive symptoms were the most common specific symptom addressed (in 9 studies), followed by anxiety symptoms (in 6 studies). Insomnia and stress symptoms were infrequently targeted and were only included in studies with multiple outcomes. Among the studies, 9 explicitly provided relevant details or numbers; no intervention had a retention rate lower than 70%, with 7 noting rates above 90%.

Systematic Review

Several interventions reported multiple follow-ups during the course of their evaluation. The range of follow-up periods varied from same day postoperation [43] to 12 months [49]. A number of different scales were used across studies to measure the internalizing disorder outcomes. In fact, the only scales used multiple times were the Beck Depression Inventory-II, Depression Anxiety Stress Scale, and Center for Epidemiological Studies–Depression Child. Most studies were conducted in a *real-world* setting, such as at home or in school locations. Other contexts included emergency rooms [44,45], medical clinics [50] or hospitals [43], and commuting settings [28]. A variety of different program delivery methods (eg, apps, monitoring programs, text messages, videos) were reported among the included studies. Only 3 studies [28,43,49] did not examine populations particularly affected by or at risk of internalizing disorders (eg, *at risk* diagnosed disorder, history of self-harm, etc).

Furthermore, 8 of the 12 studies reported an empirically proven treatment or guideline on which their intervention was based. Cognitive behavioral therapy (CBT) was the most commonly reported treatment, with 5 studies noting its use. Other treatments listed were attention bias modification [27], emotional self-awareness [42], and positive psychology, social interaction, and gameful design [50]. Among the CBT-centered interventions, specific features included concepts such as emotional regulation [45], thought modification [44], photo libraries of positive memories and physical activities [47], video diary messages [49], and psychoeducation [48]. Documented intervention delivery strategies in the non-CBT studies included visual relaxation narratives [28], smartphone app games [43], stress monitoring [46], attention bias modification task completion exercises [27], daily reporting of moods, substance use, sleep, and activities [42], symptom frequency and severity tracking [50], and videos and positive messaging [29].

Effectiveness of Interventions (Meta-Analysis)

The meta-analysis examined smartphone-based interventions as 1 total group and based on primary outcome and intervention style. Table 2 shows the results of the overall and subgroup meta-analyses. The pooled unweighted sample effect sizes of all studies (κ=12; N=1370) reflect a small-to-moderate effect (d=0.40). When sample weights are added to calculate a more credible estimate, a smaller but significant mean effect size (d=0.20; 95% CI 0.02-0.38) was observed. The sampling error variance explains approximately 3.5% of the variance in this corrected estimate. The subgroup analysis of intervention studies primarily assessing anxiety symptoms (κ=6; n=322) had a greater sample-weighted standardized effects mean compared with the overall group and was statistically significant (d=0.42; 95% CI 0.00-0.83). Among a larger sample, interventions targeting depressive symptoms (κ=9; n=1102) had a notably smaller effect that was similarly significant (d=0.16; 95% CI 0.01-0.31). When analyzing the interventions based on their predominant style and features, the sample-weighted mean effect was significant and greater for the group of all other styles (eg, monitoring, relaxation, support, not CBT) of program delivery (κ=7; n=380; d=0.42; 95% CI 0.09-0.75). Interventions using CBT features had a very small weighted mean effect size (κ=5; n=990; d=0.11; 95% CI −0.06 to 0.28) but were nonsignificant.

Table 2. Effectiveness of smartphone-based interventions on reducing youth internalizing disorder symptoms, overall, by primary targeted outcome, and by intervention style.

Analysis	κ[a]	N[b]	☒	g[d]	☒	SD_{wg}[f]	S_{*g}[g]	Variance sampling[h] (%)	95% CI[i]
Total	12	1370	114	0.40	0.20	0.07	0.11	0.04	0.02 to 0.38
Symptom									
Anxiety[j]	6	322	54	0.43	0.42	0.24	0.32	0.08	0.00 to 0.83
Depression[j]	9	1102	122	0.28	0.16	0.03	0.06	0.03	0.01 to 0.31
Intervention style									
CBT[k]-based	5	990	198	0.19	0.11	0.03	0.05	0.02	−0.06 to 0.28
Other (eg, monitoring, support, relaxation)	7	380	54	0.55	0.42	0.12	0.20	0.08	0.09 to 0.75

[a] κ: number of studies.

[b] N: total sample size for studies combined.

[c] ☒: average sample for studies combined.

[d] g: unweighted mean Hedges g.

[e] ☒: sample-weighted mean effect size (ie, weighted average of Hedges g).

[f] SD_{wg}: SD of the sampled-weighted effects.

[g] S_{*g}: corrected weighted variance of the effect (Hedges g).

[h] Percentage of variance sampling: percentage of variance because of sampling error (ie, average sampling error variance).

[i] 95% CI: CIs at 95% for sample-weighted mean.

[j] Studies with multiple outcomes were included in the subgroup analysis of both anxiety and depression.

[k] CBT: cognitive behavioral therapy.

Discussion

Principal Findings

This meta-analysis reviewed the features and explored the effectiveness of smartphone-based interventions aimed at addressing youth internalizing disorders. The review findings indicated that the majority of interventions were grounded in proven treatments, that depressive and anxiety symptoms were the predominant outcomes measured and evaluated, and that all studies were conducted in developed nations. Across the 12 studies and 1370 participants included in the meta-analyses, and after being corrected for sample error variance, a small but significant pooled effect was observed with regard to reducing internalizing disorder symptoms. When examining subgroups based on specific symptoms and intervention styles, potentially because of small κ values, mixed results were observed among the 4 analyses.

Findings from this review indicate that research conducted on the effectiveness of smartphone-based interventions for youth internalizing disorders appears to be in the nascent stages but is growing. A global review assessing the overall state of mobile health apps documented 3673 mobile phone apps but found that only 247 corresponded with published papers [53]. However, although evaluative studies are still being conducted, this particular body of literature may be growing. A similar review conducted 6 years ago documented smartphone app interventions for depression among all populations and synthesized only 8 papers [54]; this review synthesized 12 studies for 1 specific demographic. Given that during our full-text assessments, we discarded over 15 potential papers for being in the pre-evaluation phase (ie, they were one of a protocol, feasibility, or acceptability manuscript), more evaluations on this topic appear likely in the coming years.

Our review also documented that unlike findings in other related health fields, the majority of studies on this topic have been linked to proven treatments. Researchers studying smoking cessation, for instance, found that less than a third of web-based interventions were linked to proven treatment or guidelines [55]. Others examining depression apps [56], for instance, have described similar trends. Similarly, an examination of the National Health Service app library in the United Kingdom revealed that only 4 of 27 apps provided any form of evidence related to patient-reported outcomes in support of their claims of effectiveness [57], which has subsequently raised concerns about the reliability of these tools as support mechanisms. This review found that 66% (8/12) studies linked their intervention to a proven treatment, which represents an encouraging development in this regard and a chance to reiterate that such research practices should be applied in all appropriate future work.

The results of the meta-analysis confirm the tendency of digital public health interventions to produce modest-to-low effect sizes. Similar trends were reported in a systematic review and meta-analysis of internet-based promotions for health behavior change by Webb et al [58], who concluded that such "interventions had a statistically small but significant effect on health-related behavior." Typically, these outcomes are argued to be a result of a myriad of confounding variables (eg, genetic predispositions, culture, seasonality); however, the mixed effectiveness of interventions targeting youth populations has previously been considered to derive from poor or insufficient efforts to adapt the initiatives to children's cognitive and developmental needs [59]. A lack of professional support, tailoring to individual needs, and formulation taking into account immediate family and school contexts have been specifically noted [60]. Given that tailoring or personalization of public health interventions has been found to be desirable [61], it may be worthwhile for future interventions dealing with younger populations to incorporate such strategies, especially those pertaining to cognitive abilities and app support. In fact, in a study of Australian adolescents who were enrolled in a digital intervention, participants professed to value the program's relatability, narrative structure, and personal choice aspects [62]. Interestingly, however, across the reviewed studies in this review there were, generally, relatively high retention rates—unlike earlier reviews on similar topics such as computer-tailored intervention for behavior change (74.4%) [63]—although, this is likely because of many small-scale evaluations. Nevertheless, an enhanced emphasis on these aspects of program delivery going forward may help improve the effectiveness of smartphone-based interventions addressing youth populations.

Cross-study assessments of specific internalizing disorder symptoms (eg, anxiety) and intervention style (eg, CBT based) in our subgroup analyses revealed mixed significant effects. Interestingly, studies that included interventions featuring CBT, which is a therapeutic approach that has been recommended for the treatment of mental health issues among adolescents [64], only produced a small, nonsignificant pooled effect size. This may possibly be a consequence of such interventions still being in their nascent stages and having only been tested on limited populations. For comparison, Webb et al [58] similarly observed small effects in their overall analyses of internet-based interventions; they also found no significant effects for smaller subgroup analyses (eg, smoking abstinence, model or demonstration behavior change techniques). There are a few other possible reasons for this outcome. First, this was likely a consequence of the Whittaker et al [49] article's relatively large sample and correspondingly small effect exerting its influence over the other smaller studies. Second, known barriers to CBT include a lack of training, infrastructure, and funding [65], which may have been present in some studies. Finally, our conservative approach, specifically the decision to use the last follow-up reports and not the most recent postintervention reports resulted in the smallest calculations of Hedges g values (effect sizes) used in the final meta-analysis.

Taking the review and meta-analysis results together, future studies are warranted to better understand the specific impacts of smartphone-based interventions on different internalizing disorder symptoms as well as their effectiveness as a public health program delivery method. Noting the suggested potential of theory-informed interventions from the review of Webb et al [58] and the small but significant effects found in this review, it would be prudent for future interventions targeting youth mental to continue including proven treatments and pair them with behavioral or social change theories in their delivery methods. A similarly important area will be developing robust evaluation frameworks. As we identified varying lengths of follow-up evaluations, a diverse range of assessment scales, and a variety of different delivery strategies (eg, text messages, apps, monitoring) among the included studies, developing methods to precisely understand and evaluate smartphone-based interventions for sustainability, efficacy or user satisfaction, and functionality should be a priority. On this point, Chan et al [66] have previously recommended that apps be evaluated based on 3 central criteria: integration or infrastructure, usability, and usefulness. Future evaluative frameworks may also consider criteria related to support in the form of self-help strategies, as the method has the potential to relieve some of the burden on existing health care services [67] and has provided positive results for mental health interventions in individual research studies [68].

Limitations

There are important caveats to note when interpreting the results of this meta-analysis. As an area of study, research on smartphone-based interventions targeting youth and young people's mental health, specifically internalizing disorders, is still emerging when compared with other areas of public mental health research. Many of the corrected estimates presented in this meta-analysis were thus derived from small sample populations and a limited number of studies overall. In addition, owing to these small numbers, we did not correct for any other potential errors such as attenuation or dichotomization in the meta-analysis. Such analytical limitations are important to disclose given that the effects of public health interventions are, as previously noted, typically confounded by several variables that may not be captured in the measurement of the primary outcomes and additional corrections could enhance the insights of a meta-analysis. Similar to other recent meta-analyses of health behavior change interventions [69], our risk of bias assessments varied widely across evaluated studies and their intervention designs and should be considered in the context of potential assessor biases being present. Finally, several different scales and measures were used to assess internalizing disorders, potentially resulting in some discrepancy in the measurements of the outcomes used in the analysis. Although the observed between-study variance because of sampling noted in our results was rather modest, which is likely because of the overpowering impact of the Whittaker et al [49] sample and age restrictions of this review—there is likely a high level of variation in our findings derived from the clinical and methodological heterogeneity (eg, varying levels of randomization, types of interventions, controls in analyses) of included studies. On the basis of this, we encourage more specifically focused future meta-analyses that contain a greater number of studies to assess these forms of heterogeneity in their analyses.

Conclusions and Future Recommendations

Smartphone-based interventions targeting youth populations appear to be an efficacious strategy to address symptoms of internalizing disorders. This systematic review and meta-analysis found small but significant pooled effects sizes for smartphone-based interventions in reducing the symptoms of internalizing disorders among youth. However, the results also clarify the need for more research in this area. More empirical research studies conducted on a wider range of populations and settings and development of evaluative frameworks for smartphone-based intervention are recommended for future study. Furthermore, our meta-analysis confirms that only a few of the identified conference proceedings, feasibility studies, and other reports have been comprehensively and rigorously evaluated. By following these suggestions, it is possible to further improve not only the understanding of the impact of smartphone-based interventions on youth populations but also better assess the efficacy of smartphones as a mechanism of change for youth internalizing disorders.

Conflicts of Interest

None declared.

Multimedia Appendix 1
General characteristic of the included studies.
[DOCX File , 17 KB - jmir_v23i1e16490_app1.docx]

Multimedia Appendix 2
Risk of bias in randomized studies.
[PNG File , 314 KB - jmir_v23i1e16490_app2.png]

Multimedia Appendix 3
Risk of bias in nonrandomized studies.
[PNG File , 428 KB - jmir_v23i1e16490_app3.png]

References

1. Improving the mental and brain health of children and adolescents. World Health Organization. URL: http://www.who.int/mental_health/maternal-child/child_adolescent/en/ [accessed 2019-06-11]
2. The Mental Health Foundation. Childhood and adolescent mental health: understanding the lifetime impacts. Office of Health Economics. London, UK; 2005 Jan. URL: https://www.mentalhealth.org.uk/sites/default/files/lifetime_impacts.pdf [accessed 2019-07-04]
3. Haller DM, Sanci LA, Sawyer SM, Patton GC. The identification of young people's emotional distress: a study in primary care. Br J Gen Pract 2009 Mar;59(560):e61-e70 [FREE Full text] [doi: 10.3399/bjgp09X419510] [Medline: 19275825]
4. Suffoletto B, Wagner AK, Arenth PM, Calabria J, Kingsley E, Kristan J, et al. Mobile phone text messaging to assess symptoms after mild traumatic brain injury and provide self-care support: a pilot study. J Head Trauma Rehabil 2013;28(4):302-312. [doi: 10.1097/HTR.0b013e3182847468] [Medline: 23474882]
5. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
6. Cosgrove VE, Rhee SH, Gelhorn HL, Boeldt D, Corley RC, Ehringer MA, et al. Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. J Abnorm Child Psychol 2010 Aug 4;39(1):109-123. [doi: 10.1007/s10802-010-9444-8]
7. Krain A, Ghaffari M, Freeman J, Garcia A, Leonard H, Pine D. Lewis's Child Adolescent Psychiatry. In: Lewis's Child and Adolescent Psychiatry: A Comprehensive Textbook. Philadelphia, Pennsylvania: Lippincott Williams and Wilkins; 2007:538-547.
8. Hazell P. Depression in children and adolescents. BMJ Clin Evid 2009 Jan 07;2009:- [FREE Full text] [Medline: 19445770]
9. Depression. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2019-07-06]
10. Cairney J, Rigoli D, Piek J. Developmental coordination disorder and internalizing problems in children: the environmental stress hypothesis elaborated. Dev Rev 2013 Sep;33(3):224-238. [doi: 10.1016/j.dr.2013.07.002]
11. Gregory AM, Buysse DJ, Willis TA, Rijsdijk FV, Maughan B, Rowe R, et al. Associations between sleep quality and anxiety and depression symptoms in a sample of young adult twins and siblings. J Psychosom Res 2011 Oct;71(4):250-255. [doi: 10.1016/j.jpsychores.2011.03.011] [Medline: 21911103]
12. Madrid-Valero J, Ronald A, Shakeshaft N, Schofield K, Malanchini M, Gregory A. Sleep quality, insomnia, and internalizing difficulties in adolescents: insights from a twin study. Sleep 2020 Feb 13;43(2):-. [doi: 10.1093/sleep/zsz229] [Medline: 31553447]
13. Whalen DJ, Sylvester CM, Luby JL. Depression and anxiety in preschoolers: a review of the past 7 years. Child Adolesc Psychiatr Clin N Am 2017 Jul;26(3):503-522 [FREE Full text] [doi: 10.1016/j.chc.2017.02.006] [Medline: 28577606]

14. Rapee RM, Schniering CA, Hudson JL. Anxiety disorders during childhood and adolescence: origins and treatment. Annu Rev Clin Psychol 2009;5:311-341. [doi: 10.1146/annurev.clinpsy.032408.153628] [Medline: 19152496]
15. Compton SN, Burns BJ, Helen LE, Robertson E. Review of the evidence base for treatment of childhood psychopathology: internalizing disorders. J Consult Clin Psychol 2002 Dec;70(6):1240-1266. [doi: 10.1037//0022-006x.70.6.1240] [Medline: 12472300]
16. Konnopka A, Leichsenring F, Leibing E, König HH. Cost-of-illness studies and cost-effectiveness analyses in anxiety disorders: a systematic review. J Affect Disord 2009 Apr;114(1-3):14-31. [doi: 10.1016/j.jad.2008.07.014] [Medline: 18768222]
17. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord 2018 Jan 23;6(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0110-8] [Medline: 29359252]
18. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela J. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. npj Digital Med 2018 Apr 6;1(1):1-9. [doi: 10.1038/s41746-018-0022-8] [Medline: 28688902]
19. Mehrotra S, Tripathi R. Recent developments in the use of smartphone interventions for mental health. Curr Opin Psychiatry 2018 Sep;31(5):379-388. [doi: 10.1097/YCO.0000000000000439] [Medline: 29985179]
20. Oh SJ, Seo S, Lee JH, Song MJ, Shin M. Effects of smartphone-based memory training for older adults with subjective memory complaints: a randomized controlled trial. Aging Ment Health 2018 Apr;22(4):526-534. [doi: 10.1080/13607863.2016.1274373] [Medline: 28071929]
21. Moëll B, Kollberg L, Nasri B, Lindefors N, Kaldo V. Living SMART — A randomized controlled trial of a guided online course teaching adults with ADHD or sub-clinical ADHD to use smartphones to structure their everyday life. Internet Interv 2015 Mar;2(1):24-31. [doi: 10.1016/j.invent.2014.11.004]
22. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
23. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
24. Domhardt M, Steubl L, Baumeister H. Internet- and mobile-based interventions for mental and somatic conditions in children and adolescents. Z Kinder Jugendpsychiatr Psychother 2020 Jan;48(1):33-46 [FREE Full text] [doi: 10.1024/1422-4917/a000625] [Medline: 30422059]
25. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
26. Petticrew M, Roberts H. Systematic Reviews in the Social Sciences: A Practical Guide. Malden, MA, USA: Blackwell Pub; 2006:1-335.
27. Clarke PJ, Bedford K, Notebaert L, Bucks RS, Rudaizky D, Milkins BC, et al. Assessing the therapeutic potential of targeted attentional bias modification for insomnia using smartphone delivery. Psychother Psychosom 2016;85(3):187-189. [doi: 10.1159/000442025] [Medline: 27043732]
28. Grassi A, Gaggioli A, Riva G. The green valley: the use of mobile narratives for reducing stress in commuters. Cyberpsychol Behav 2009 Apr;12(2):155-161. [doi: 10.1089/cpb.2008.0156] [Medline: 19275475]
29. Takahashi K, Takada K, Hirao K. Feasibility and preliminary efficacy of a smartphone application intervention for subthreshold depression. Early Interv Psychiatry 2019 Feb;13(1):133-136. [doi: 10.1111/eip.12540] [Medline: 29356332]
30. Keith DJ, Rinchuse DJ, Kennedy M, Zullo T. Effect of text message follow-up on patient's self-reported level of pain and anxiety. Angle Orthod 2013 Jul;83(4):605-610. [doi: 10.2319/091812-742.1] [Medline: 23210546]
31. Wilkinson P. Conceptualization about internalizing problems in children and adolescents. Cien Saude Colet 2009;14(2):373-381 [FREE Full text] [doi: 10.1590/s1413-81232009000200007] [Medline: 19197413]
32. Krueger RF. The structure of common mental disorders. Arch Gen Psychiatry 1999 Oct;56(10):921-926. [doi: 10.1001/archpsyc.56.10.921] [Medline: 10530634]
33. Watson D, Clark LA, Carey G. Positive and negative affectivity and their relation to anxiety and depressive disorders. J Abnorm Psychol 1988 Aug;97(3):346-353. [doi: 10.1037//0021-843x.97.3.346] [Medline: 3192830]
34. Hunter J, Schmidt F. Methods of Meta-analysis: Correcting Error and Bias in Research Findings. Thousand Oaks, CA, USA: Sage; 1990.
35. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
36. Allen TD, Eby LT, Poteet ML, Lentz E, Lima L. Career benefits associated with mentoring for protégeé: a meta-analysis. J Appl Psychol 2004 Feb;89(1):127-136. [doi: 10.1037/0021-9010.89.1.127] [Medline: 14769125]
37. Field AP. Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. Psychol Methods 2001 Jun;6(2):161-180. [doi: 10.1037/1082-989x.6.2.161] [Medline: 11411440]

38. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002 Mar;7(1):105-125. [doi: 10.1037/1082-989x.7.1.105] [Medline: 11928886]
39. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. URL: https://handbook-5-1.cochrane.org/ [accessed 2020-12-24]
40. Quality Assessment Tool For Quantitative Studies. Effective Public Health Practice Project. 2017. URL: https://merst.ca/wp-content/uploads/2018/02/quality-assessment-tool_2010.pdf [accessed 2018-11-29]
41. Quality Assessment Tool for Quantitative Studies Dictionary. Effective Public Health Practice Project. 2017. URL: https://merst.ca/wp-content/uploads/2018/02/qualilty-assessment-dictionary_2017.pdf [accessed 2018-11-29]
42. Kauer SD, Reid SC, Crooke AHD, Khor A, Hearps SJC, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
43. Lee J, Jung H, Lee G, Kim H, Park S, Woo S. Effect of behavioral intervention using smartphone application for preoperative anxiety in pediatric patients. Korean J Anesthesiol 2013 Dec;65(6):508-518 [FREE Full text] [doi: 10.4097/kjae.2013.65.6.508] [Medline: 24427456]
44. Ranney ML, Freeman JR, Connell G, Spirito A, Boyer E, Walton M, et al. A depression prevention intervention for adolescents in the emergency department. J Adolesc Health 2016 Oct;59(4):401-410 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.04.008] [Medline: 27267141]
45. Ranney M, Pittman S, Dunsiger S, Guthrie K, Spirito A, Boyer E, et al. Emergency department text messaging for adolescent violence and depression prevention: a pilot randomized controlled trial. Psychol Serv 2018 Nov;15(4):419-428 [FREE Full text] [doi: 10.1037/ser0000193] [Medline: 30382737]
46. Reid SC, Kauer SD, Hearps SJC, Crooke AHD, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011 Nov 29;12:131 [FREE Full text] [doi: 10.1186/1471-2296-12-131] [Medline: 22123031]
47. Stallard P, Porter J, Grist R. A smartphone app (BlueIce) for young people who self-harm: open phase 1 pre-post trial. JMIR Mhealth Uhealth 2018 Jan 30;6(1):e32 [FREE Full text] [doi: 10.2196/mhealth.8917] [Medline: 29382632]
48. Werner-Seidler A, Wong Q, Johnston L, O'Dea B, Torok M, Christensen H. Pilot evaluation of the Sleep Ninja: a smartphone application for adolescent insomnia symptoms. BMJ Open 2019 May 27;9(5):e026502 [FREE Full text] [doi: 10.1136/bmjopen-2018-026502] [Medline: 31133584]
49. Whittaker R, Stasiak K, McDowell H, Doherty I, Shepherd M, Chua S, et al. MEMO: an mHealth intervention to prevent the onset of depression in adolescents: a double-blind, randomised, placebo-controlled trial. J Child Psychol Psychiatry 2017 Sep;58(9):1014-1022. [doi: 10.1111/jcpp.12753] [Medline: 28573672]
50. Worthen-Chaudhari L, McGonigal J, Logan K, Bockbrader MA, Yeates KO, Mysiw WJ. Reducing concussion symptoms among teenage youth: evaluation of a mobile health app. Brain Inj 2017;31(10):1279-1286 [FREE Full text] [doi: 10.1080/02699052.2017.1332388] [Medline: 28665690]
51. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
52. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
53. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
54. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
55. Bock BC, Graham AL, Whiteley JA, Stoddard JL. A review of web-assisted tobacco interventions (WATIs). J Med Internet Res 2008 Nov 06;10(5):e39 [FREE Full text] [doi: 10.2196/jmir.989] [Medline: 19000979]
56. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
57. Leigh S, Flatt S. App-based psychological interventions: friend or foe? Evid Based Ment Health 2015 Nov;18(4):97-99. [doi: 10.1136/eb-2015-102203] [Medline: 26459466]
58. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
59. Adelman CB, Panza KE, Bartley CA, Bontempo A, Bloch MH. A meta-analysis of computerized cognitive-behavioral therapy for the treatment of DSM-5 anxiety disorders. J Clin Psychiatry 2014 Jul;75(7):e695-e704. [doi: 10.4088/JCP.13r08894] [Medline: 25093485]

60. Simons L, Valentine AZ, Falconer CJ, Groom M, Daley D, Craven MP, et al. Developing mHealth remote monitoring technology for attention deficit hyperactivity disorder: a qualitative study eliciting user priorities and needs. JMIR Mhealth Uhealth 2016 Mar 23;4(1):e31 [FREE Full text] [doi: 10.2196/mhealth.5009] [Medline: 27009498]
61. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
62. Cheek C, Bridgman H, Fleming T, Cummings E, Ellis L, Lucassen MF, et al. Views of young people in rural Australia on SPARX, a fantasy world developed for New Zealand youth with depression. JMIR Serious Games 2014 Feb 18;2(1):e3 [FREE Full text] [doi: 10.2196/games.3183] [Medline: 25659116]
63. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
64. Connolly SD, Bernstein GA, Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with anxiety disorders. J Am Acad Child Adolesc Psychiatry 2007 Feb;46(2):267-283. [doi: 10.1097/01.chi.0000246070.23695.06] [Medline: 17242630]
65. Gunter RW, Whittal ML. Dissemination of cognitive-behavioral treatments for anxiety disorders: overcoming barriers and improving patient access. Clin Psychol Rev 2010 Mar;30(2):194-202. [doi: 10.1016/j.cpr.2009.11.001] [Medline: 19942331]
66. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
67. Jorm AF, Griffiths KM. Population promotion of informal self-help strategies for early intervention against depression and anxiety. Psychol Med 2006 Jan;36(1):3-6. [doi: 10.1017/S0033291705005659] [Medline: 16356291]
68. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. Br Med J 2012 Apr 19;344(apr18 3):e2598-e2598. [doi: 10.1136/bmj.e2598]
69. Maher CA, Lewis LK, Ferrar K, Marshall S, De BI, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]

Abbreviations

CBT: cognitive behavioral therapy
EPHPP: Effective Public Health Practice Project
QA: quality assessment
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 03.10.19; peer-reviewed by M Domhardt, J Bender; comments to author 26.10.20; revised version received 11.11.20; accepted 18.11.20; published 11.01.21.

Please cite as:
Buttazzoni A, Brar K, Minaker L
Smartphone-Based Interventions and Internalizing Disorders in Youth: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(1):e16490
URL: http://www.jmir.org/2021/1/e16490/
doi:10.2196/16490
PMID:33427682

Review

How the Use of a Patient-Accessible Health Record Contributes to Patient-Centered Care: Scoping Review

Janine Benjamins[1,2], MD; Annemien Haveman-Nies[2,3], PhD; Marian Gunnink[1], MD; Annemieke Goudkuil[1], MSc; Emely de Vet[2], PhD

[1]Icare JGZ, Meppel, Netherlands

[2]Chairgroup Consumption & Healthy Lifestyles, Wageningen University, Wageningen, Netherlands

[3]GGD NOG, Warnsveld, Netherlands

Corresponding Author:
Janine Benjamins, MD
Icare JGZ
Blankenstein 550
Meppel, 7943 PH
Netherlands
Phone: 31 612329494
Email: j.benjamins@cjgnoordveluwe.nl

Abstract

Background: Worldwide, patient-centered care is becoming a widely used concept in medical practice, getting more and more attention because of its proven ability to improve quality of care and reduce costs. Although several studies show that patient-accessible electronic health records (PAEHRs) influence certain aspects of patient-centered care, the possible contribution of PAEHR implementation to patient-centered care as a comprehensive concept has not, to our knowledge, been structurally evaluated to date.

Objective: The objective of this study is to review whether and how the use of PAEHRs contributes to patient-centered care both in general and among specific population groups.

Methods: We followed PRISMA Extension for Scoping Reviews reporting guidelines. We identified literature in 5 databases, using the terms "patient-accessible medical records," "patient experiences," and "professional experiences" as key concepts. A total of 49 articles were included and analyzed with a charting code list containing 10 elements of patient-centered care.

Results: Studies were diverse in design, country of origin, functionalities of the investigated PAEHR, and target population. Participants in all studies were adults. Most studies reported positive influence of PAEHR use on patient-centered care; patient accessible health records were appreciated for their opportunity to empower patients, inform patients about their health, and involve patients in their own care. There were mixed results for the extent to which PAEHRs affected the relation between patients and clinicians. Professionals and patients in mental health care held opposing views concerning the impact of transparency, where professionals appeared more worried about potential negative impact of PAEHRs on the patient-clinician relationship. Their worries seemed to be influenced by a reluctant attitude toward patient-centered care. Disadvantaged groups appeared to have less access to and make less use of patient-accessible records than the average population but experienced more benefits than the average population when they actually used PAEHRs.

Conclusions: The review indicates that PAEHRs bear the potential to positively contribute to patient-centered care. However, concerns from professionals about the impact of transparency on the patient-clinician relationship as well as the importance of a patient-centered attitude need to be addressed. Potentially significant benefits for disadvantaged groups will be achieved only through easily accessible and user-friendly PAEHRs.

(J Med Internet Res 2021;23(1):e17655) doi:10.2196/17655

KEYWORDS

personal health records; patient portals; patient-centered care; patient-accessible records

Introduction

In the last 30 years, patient-centeredness has grown worldwide in relevance in health care policy, practice, and research. In 1987, Harvey Picker developed the Pickers' Principles of Patient Centered Care [1]. Thereafter, patient-centered care gained increasing prominence in the US when the Institute of Medicine advocated for patient-centered care as a cornerstone of health care quality [2]. In 2015, the World Health Organization stated that patient-centered care should become the standard for health care systems all over the world [3].

Key factors in patient-centered care are responsiveness to the patients' individual needs and preferences, and partnership between care providers and patients in decision making [4-7]. Patients are acknowledged as unique human beings with needs and preferences that have to be taken into account when clinical decisions are made. Ideally, patients as well as their family members or caregivers are involved in making these decisions. This requires clear information and communication with patients.

Patient-centered care has been gaining importance because of its proven ability to increase the quality of care, with lower health care utilization as a beneficial side effect [3,8-13]. The growing importance and development of the concept in different countries has led to a diversity in models, definitions, and terminology. For this review, we used an integrative model by Scholl et al [5], integrating more than 400 definitions and models into a new and comprehensible model for patient-centered care.

In the Netherlands, patient-centered care has also taken center stage in the discussion about quality of care, especially in care for youth [14]. To contribute to patient-centered care, three organizations for preventive youth health care and youth social services in the North Veluwe region developed a PAEHR system [14]. The assumption that the use of PAEHRs contributes to patient-centered care, however, has not yet been sufficiently proven.

Several reviewers investigated effects of PAEHRs by reporting on a variety of outcomes related to patient health, quality of care, or patient satisfaction [15-23]. The aspects of patient-centered care that have been mentioned are, for instance, empowerment of patients, trust in care providers, and the clinician-patient relationship. For these aspects, both beneficial [15-19] and unfavorable or even harmful consequences of the use of a PAEHR [19-23] to patient-centered care have been reported. Some studies report that disadvantaged groups might benefit less from the use of PAEHRs than others, as their access to and use of PAEHRs is lower than average [19,20,22,23]. To date, we know of no published review that structurally evaluates the possible contribution of PAEHRs to patient-centered care as a comprehensive concept. Performing such a review would enable us to explore whether PAEHRs could serve as a tool to strengthen this value-based health care model.

Since the relationship between the use of PAEHRs and the broad concept of patient-centered care has, to date, received limited attention in reviews, a broad overview of recent literature is required, with inclusion of different study designs. With such a broad perspective, a scoping review is more suitable than a systematic review, as scoping reviews aim to broadly summarize and synthesize evidence instead of finding answers to circumscript questions and including only specified study designs. A scoping review can be helpful to provide direction to future research and search for gaps in knowledge [24,25]. The objective of this review is to provide an overview of recent literature about experiences of patients and professionals with the use of PAEHRs and to investigate whether and how the use of PAEHRs contributes to patient-centered care, both in general and among specific population groups.

Methods

Search Strategy and Inclusion Criteria

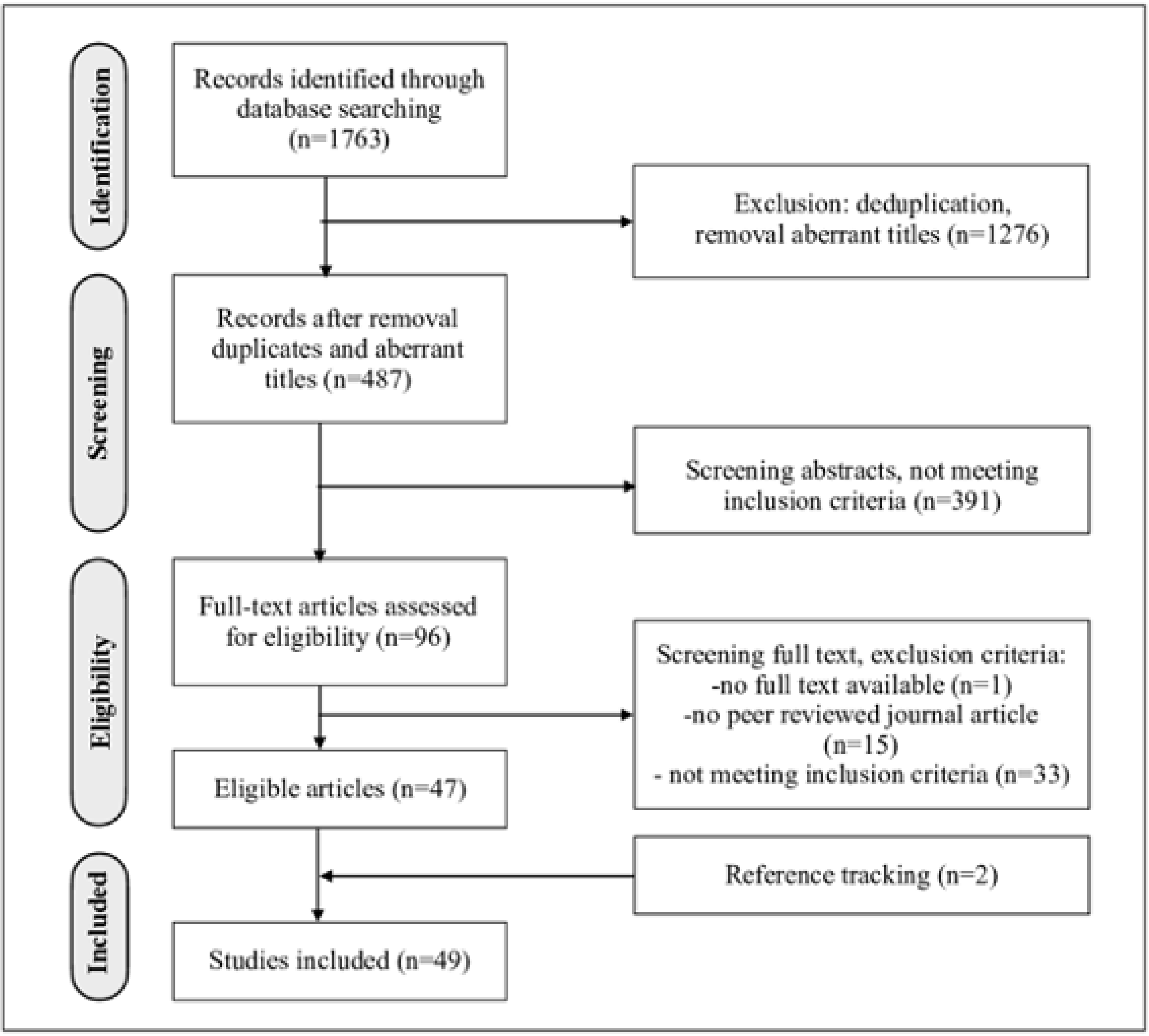
Design and reporting of this scoping review were in line with the framework for scoping reviews by Arksey and O'Malley [24-26], which was further developed by other authors, finally leading to the PRISMA Extension for Scoping Reviews guideline and checklist [27,28]. Multimedia Appendix 1 contains the completed PRISMA checklist for this review. The a priori review protocol has not been registered. Key concepts used in the search were "patient-accessible medical records," "patient experiences," and "professional experiences." Table 1 contains the full electronic search string for the Scopus database. The search was limited to papers written in English or Dutch, being languages all authors understand, and to studies published between January 2000 and April 2019. This period was chosen because, in a first quick search, most articles about PAEHRs appeared to originate from 2000 or more recently. Five databases were searched: (1) Pubmed, (2) Medline, (3) Scopus, (4) Socindex, and (5) Psychinfo. The final search was run on April 9, 2019. Search records were uploaded to Endnote X8 to facilitate the article selection process.

Searches, deduplication, and first screening of titles were performed by SJB. In total, 1763 articles were found and screened for eligibility (Figure 1). Aberrant titles were removed, and abstracts of remaining articles were independently screened by different individuals (SJB, MG, and AG), in line with the scoping nature of the review. We included research articles from peer reviewed journals for which full text could be retrieved. The articles were based on original research data. They addressed "experiences" of professionals or patients/clients using a PAEHR. Articles were screened in 3 rounds. After every round, different interpretations were discussed between all three screening authors to come to a unanimous decision. If necessary, the inclusion criteria were adapted before the next round to ensure uniform selection. SJB screened the remaining full text articles on inclusion criteria. To exclude articles from predatory journals, every journal was checked against the JournalGuide whitelist [29]. The selection process was finalized by reference tracking; all references of selected articles were checked with the inclusion criteria and added when eligible.

Table 1. Full search string for Scopus, split into three key concepts.

Key concepts	Search string per concept
Patient-accessible	(“Patient” OR “Patients” OR “client” OR “clients”) AND (“access” OR “online access” OR “accessible”) AND (“record” OR “records” OR “file” OR “files”)
Medical records	AND “Personal health records” OR “Health Record, Personal” OR “Personal Health Record” OR “Record, personal health” OR “personal health records” OR “Personal Health information” OR “Health Information, Personal” OR “Information, Personal Health” OR “Personal Medical Records” OR “Medical Record, Personal” OR “Medical Records, Personal” OR “Personal Medical Record” OR “Record, Personal Medical” OR “Records, Personal Medical” OR “patient portals” OR “Patient Web Portal” OR “Portal, Patient Web” OR “Portals, Patient Web” OR “Web Portal, Patient” OR “Web Portals, Patient” OR “Patient Internet Portals” OR “Internet Portal, Patient” OR “Internet Portals, Patient” OR “Patient Internet Portal” OR “Portal, Patient Internet” OR “Portals, Patient Internet” OR “Patient Web Portals” OR “Patient Portal” OR “Portal, Patient” OR “Open Notes” OR “Electronic health records”
Patient experiences AND physician experiences	AND “patient experiences” OR “physician experiences” OR “experiences” OR “experiences, patient” OR “experiences, patients” OR “experiences, physician” OR “experiences, physicians” OR “experiences, professional” OR “professional experiences” OR “outcome assessment (health care)” OR “benefit” OR “satisfaction” OR “patient outcomes”

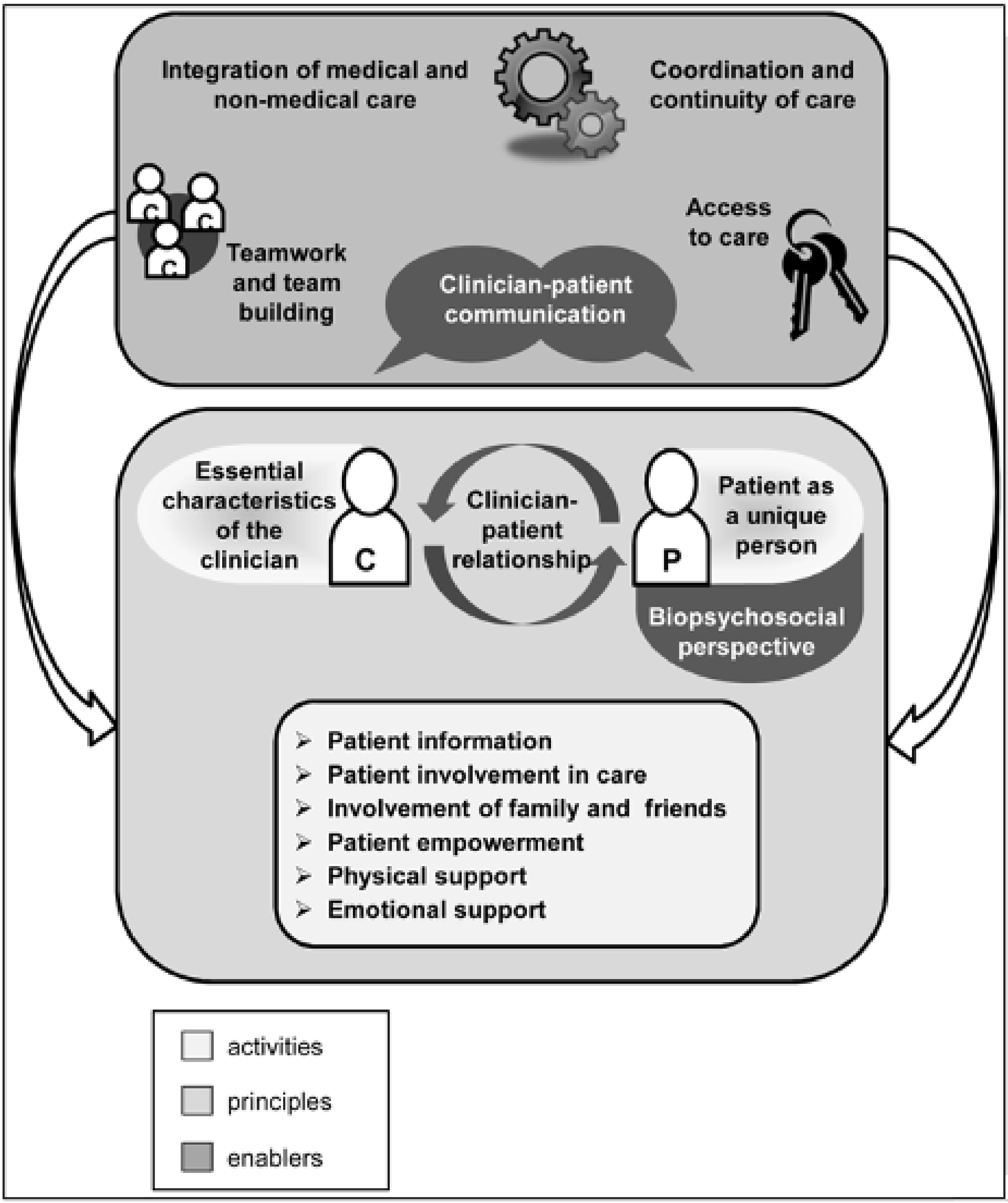
Figure 1. Flow diagram of article selection.



Data Analysis

Through discussion SJB, AH, and EV came to a charting code list (see Multimedia Appendix 2). The list contained codes for general article information, study methods, description of the investigated PAEHR, and 10 dimensions of patient-centered care. The dimensions of patient-centered care were derived from a model, developed by Scholl et al (Figure 2) [5]. This model distinguishes 15 dimensions in 3 groups: (1) principles, (2) enablers, and (3) activities. The principles represent the essential factors of a patient-centered attitude in professionals. The principles and the enablers, which are organizational conditions for patient-centeredness, lay the foundation for the last group, the activities. These are actions and measures by which patient-centered behavior becomes visible. Assuming that use of PAEHRs would affect the "activities" from the model, possibly affect the "enablers," and not affect the "principles," we included all 5 enablers and 4 activities. We did not include the activities "physical support" and "emotional support," since we expected not to find any relation with the use of PAEHRs. From the principles, only clinician-patient relationship was included, because we considered this dimension a dynamic one that could be influenced by use of a PAEHR. A separate charting code was created for differences among population groups, since former research suggests that disadvantaged groups might benefit less from the use of PAEHRs than others [19,20,22,23]. The charting process was done by SJB and discussed afterward with the other authors. All charted data were aggregated through group discussion with all co-authors.

Figure 2. Model of Patient-centered Care, by Scholl et al (2014).



Results

Overview

In total, 49 eligible articles were included [21,30-77]. Multimedia Appendix 3 presents a brief summary of the articles, with characteristics of each study, functionalities of the studied PAEHR, and reported elements of patient-centered care. Multimedia Appendix 4 provides an overview of all outcomes. In this appendix, the articles were divided into 3 study design groups to facilitate the analysis. The largest group (n=34) consists of descriptive studies, both qualitative and quantitative [21,32-35,37-43,45,47-51,53,55,56,58,59,62,63,65,69-73, 75-77]. The other 2 groups contain pre-post-test comparative studies [21,40,60,61, 70,71,75,76] and studies comparing intervention and control groups [30,31,36,44,46,47,52, 54,57,64,66-68,74]. The results of 7 mixed methods studies were divided and categorized according to the groups they best matched with [21,40,47,70,71,75,76].

Most articles (n=29) originated from the US [21,30,32-40,42,44-46,49,50,54-58,60,63,66,69,74,76,77]. Clustered in 5-year periods, 3 articles originated from 2000-2004 [34,40,66], 3 from 2005-2009 [43,67,68], 15 from 2010-2014 [21,33,47,49,54,55,58,59,64,69,71,73-75,77], and 28 from 2014-2019 [30-32,35-39,41,42,44-46,48,50-53,56,57,60-63, 65,70,72,76]. Duration of experience with a PAEHR varied from 1.5 to 48 months. Population sizes were also diverse, ranging from 9 in a qualitative study [41] to several thousand in an Open Notes survey study (n=29,191) [56]. Finally, the population demographics varied; most studies included a broad range of patients (eg, patients in hospitals [30,34,60,72,76] or in primary care [21,32,42-46,48,49,54,57,58,63]). Other studies focused on specific patient groups, like cancer patients [30,37,50,59,62,75], cardiac patients [40,66,74], chronically ill patients [31,64,71], HIV-positive patients [36,57], psychiatric patients [35,39,70], gynecologic patients [67,68,73], and veterans [35,36,39,55,56,69,77]. Ten studies investigated experiences of both patients and their care providers [21,32,34,40,51,57,58,60,70,72]. Four studies focused on professionals only [38,41,61,65]. Respondents in all studies were adults, mostly of no specific age group. Three studies surveyed parents of pediatric patients [33,37,51].

Apart from record-access, the most common functionalities of the PAEHRs were “electronic messaging” [33,36,37, 40,51,55,56,64-66,68-70,72,74,76] and the possibility to add or edit health information [31,34,54-56,58,62,65,70,72,74,75]. Six studies investigated a so-called active PAEHR that sent patients “personalized health messages” [31,34,58,63,67,68]. Other functionalities were “give feedback on health information”[62,65], “download information to share with others”[30,42,58], “grant direct access to others” [55,62,76,77], and administrative tasks like “scheduling appointments” [30,51,59], “paying bills” [30], and “requesting medication refill” [30,72].

One patient-accessible record was paper-based and consisted of a briefcase with all medical information, which was updated after every visit to the clinic [47]. Two PAEHRs were electronic but not available online [43,73]. One was a USB-stick containing all medical information, which was revised during every visit to the clinic [73]. The other was a kiosk in the clinic’s waiting room, where patients could access all medical info during their visit [43]. In one study, 9 physicians were interviewed about their experiences with PAEHRs in general [30].

Dimensions of Patient Centeredness

The outcomes for the 10 coded dimensions of patient-centered care have been summarized in Table 2. In 34 of the studies at least 3 of these dimensions were explored. None of the studies mentioned the dimensions “integration of medical and non-medical care” and “teamwork”. The following paragraphs describe the outcomes for each dimension of patient-centered care. When describing outcomes, we use the term “effect” both for experienced effects as well as for objective results from comparative studies.

Table 2. Summary of results for dimensions of patient-centered care. This table represents, for every explored dimension of patient-centered care, whether reported outcomes point in a positive or negative direction. “Negative” in a pre-post comparative design means “less positive than expected.” In a pre-post or intervention-control design, the term “neutral” refers to the outcome “no difference” or “no significant difference.”

Dimension	Number of studies, n	Descriptive studies, reference number		Comparative studies, reference number		
		Positive	Negative	Positive	Neutral	Negative
Information	40	[21,32-34,37-43,45,47-51,53,55,56,58, 59,62,63,65,69-72,75-77]	N/A[a]	[30,31,36][b]; [44,46,47,76]	[60,67]	[61]
Involvement in care	33	[32,34,37,38,40-43,47-51,55,56,59,62, 63,65,69,71,75,77]	N/A	[36,54][b]; [44]	[30,46,64,67,74]	[60,61]
Empowerment	23	[21,33,39,42,45,47,48,50,56,58,63,70,76]	N/A	[46,59,60,66]; [76][b]	[36,40,61,68, 70,71,75]	N/A
Communication	22	[33,34,37,40,41,45,47,48,51,53,55,58,59, 62,63,65,70,71,76,77]	N/A	[76]	[57,66]	N/A
Involvement of family and friends	14	[42,45,47,49,51,55,59,62,69,71,73,76]; [70][c]	[70][c]	[57]	N/A	N/A
Clinician-patient relationship	22	[21,32,35,38-43,45,50,59,62,63,65,71,72]	[41][c]	[44,74]	[57]	[60,61]
Access to care	5	[42,45,49,62,63]	N/A	N/A	N/A	N/A
Coordination / continuity of care	3	[40,58,76]	N/A	N/A	N/A	N/A
Integration medical / nonmedical	0	N/A	N/A	N/A	N/A	N/A
Teamwork	0	N/A	N/A	N/A	N/A	N/A

[a] N/A: not applicable.

[b] Significant effect.

[c] Both positive and negative aspects reported.

Information

Forty studies investigated if and in what way patients felt more informed about their health after use of a PAEHR. We distinguished 3 different topics: (1) what patients valued in reading records, (2) emotional consequences, and (3) understandability. Seven descriptive studies examined reasons for reading medical records [32,43,45,56,62,75,76]. Patients valued reading their record because they wanted to know about their health or because they wanted to be sure they understood what the doctor said or because they were curious. Patients valued reading their records most because it improved understanding of health issues [21,34,39, 45-47,50,53,56,60,61,65,69,71,75-77], helped to prepare for next visits [21,56,59,61-63,65,71,75,76], and helped to remember the care plan [21,40,42,43,45,46,49,50,56,61,76]. Reading also helped patients to follow treatment recommendations [33,39,41]. Six studies compared the difference in health knowledge between intervention and control groups [30,31,36,44,47,67]. One study found a significantly higher “self-health management knowledge score” among PAEHR adopters than among nonadopters (*P*<.01) [30]. Another study found that the intervention group was significantly better informed than the control group about their latest blood measurement levels, including date, time, and trend changes, and about normal lab values (*P*<.001) [31]. A third study found that HealtheVet users were able to correctly identify their CD4 counts significantly more often (Fisher exact test=.048) and their viral load (Fisher exact test=.003) than nonusers [36]. The other studies found no significant difference [44,47,67]. Two pre-post studies compared expectations with experiences [61,76]. After a period of PAEHR use, one of the studies reported better understanding of care plans among patients than expected (OR=1.39) [76]. In the other study, however, interviewed psychiatrists reported less improvement than expected in the extent to which patients understood their medical conditions or remembered their care plans [61].

Reading their records also provided patients with reassurance [33]. In 4 qualitative studies, patients said that transparency reduced anxiety and stress [33,45,56,62]. They experienced waiting for news as more stressful than reading notes by themselves. One patient said: “It is easier to break down at home where you are surrounded by family, than at the doctor’s office” [62]. If reading records caused stress, this was in most cases related to new diagnoses which had not yet been discussed with the professional [33]. Stress was also caused if health care professionals trivialized a patient’s problem in the record [39]. Less than 10% of patients often or always experienced worries or confusion after reading their record [21,39,56,57,76]. Three intervention-control studies found no significant difference in anxiety levels or reported worries between users and nonusers [47,52,57].

Six studies investigated if patients understood everything they read and how they felt they did not understand [34,40,50,58,62,77]. Some patients said they would appreciate built-in-definitions and less jargon. On the other hand, one patient added: “I would rather have the doctors just write what they write and me work to understand it, than them writing it for me and leaving something out that I would like to know”

[40]. Moreover, although patients found some medical terminology too difficult, they managed to find explanations on the internet [58,62].

Involvement in Care

Thirty-two studies described the impact of use of PAEHRs on involvement in care. Twenty-three descriptive studies described involvement of patients in their care as a benefit of using a PAEHRs [32,34,37,38,40-43,47-51,55,56,59,62,63,65,69,71,75,77]. Clinicians in one study said that using a PAEHR resulted in a “power shift” towards patients. Some of them saw this as a “move towards patient-centered care, creating better opportunities for collaboration with patients” [38]. In intervention-control studies, the 13-question Patient Activation Measurement (PAM-13) Questionnaire was most commonly used to measure involvement of patients in their care. Two intervention-control studies found a significantly higher PAM-score in the user groups [36,54]. One study reported a mean PAM-13 score of 47 points in the intervention group versus 45 points in the control group (P=.0014) [54], whereas the other study reported a mean PAM-13 score of 72.5 in the intervention group versus a mean of 63.49 in the control group (P=.03) [36]. Three studies found no significant effect on activation score or decision making [64,67,74]. One study, comparing different user subgroups, reported that less educated patients and non-White patients were more likely to report that reading visit notes was extremely important to engaging in their care than more educated and White patients [46]. In the 2 pre-post comparisons, the observation that patients were “feeling more in control” was slightly lower than expected [60,61].

Five studies investigated if patient involvement would result in patients finding and correcting errors in their record [45,60,62,65,76]. One descriptive study reported that 6 patients in a group of 15 had found errors but had not requested correction [62]. One study investigated a PAEHR with a feedback option [45]. Patients valued this feedback option because it helped them to correct errors. Two descriptive studies reported that physicians felt that use of PAEHRs could prevent medical errors and that the PAEHRs were used by patients as a means to check for accuracy [65,76]. In one pre-post study, patients found less errors than expected, although errors were found and corrected; in a group of 50 patients, 3 patients reported finding errors in medication, 2 patients found errors in radiology test reports, and 1 patient found an error in a laboratory test report [60].

Involvement of Family and Friends

Fourteen studies investigated whether and how family and friends were involved in care through use of PAEHRs. Thirteen descriptive studies reported that patients shared health information with relatives, friends, and health professionals [42,45,47,49,51,55,59,62,69-71,73,76]. Patients said they shared information to answer questions of family and friends and to keep them informed. Sharing information also helped to discuss their disease with relatives or caregivers. The percentage of patients who actually shared notes with others differed among studies, from 15% to 67%. One descriptive study among patients with a bipolar disorder reported that 23% of the 39 respondents considered access to family caregivers preferable, whereas 25% thought it would be harmful [70]. One study, comparing HIV-positive patients with other patients in primary care, found that HIV-positive patients were more likely than other primary care patients to share or discuss visit notes with others, both friends and professionals [57]. In one mixed-methods study, caregivers especially valued the ability for a patient to share information with them, because this enabled them to view notes of visits which they had not been able to attend [76].

Empowerment

In 13 descriptive studies, patients mentioned that they felt more in control of their health or that they could take better care of their own health due to reading their record [21,33,39,42,45,47,48,50,56,58,63,70,76]. In one study, patients appreciated the possibility to share a print-out of their record with another doctor [59]. Patients also said that their role became more active [45]. They experienced more ownership of their own health status [63]. Three control-intervention studies reported no significant difference in empowerment between intervention and control groups [36,66,68]. In 7 pre-post studies, 6 studies found no significant effect on empowerment scores [40,60,61,70,71,75]. The 7th study reported that patients were more confident in their ability to manage their health information (OR 2.14, 95% CI 1.59-2.89) and their care (OR 1.48, 95% CI 1.14-1.93) [76].

Communication

Twenty descriptive studies investigated the effect on communication between patient and health care professional and reported an improvement [33,34,37,40,41,45,47,48,51,53,55,58,59,62,63,65,70,71,76,77]. Communication became easier because of the PAEHRs, and interaction improved [34,58]. The ability to view health information improved the level of communication during subsequent visits and made it possible to communicate “on a more level playing field” with health care professionals [41,51]. The use of a PAEHR also removed barriers, for instance, “because you can ask ‘stupid’ questions that you wouldn’t pick up the phone for” [33]. Two intervention-control studies reported on communication and found no significant differences between intervention and control groups [57,66]. One pre-post study reported that caregivers appreciated the possibility to view notes of visits they could not attend, because it improved their communication with care providers [76].

Seven descriptive studies investigated the influence of PAEHR use on time investment, 5 of them reporting no difference [21,32,40,58,62,65,72]. One study reported that some professionals needed more time to edit or explain notes. However, they framed this as “better documentation, a good thing” [21]. In one study, a professional said that it was improving efficiency: “finally something to save me time!” [58]. One intervention-control study reported that professionals received more messages per patient, but nonetheless did not feel a perceptible change in workload [66]. Four pre-post studies investigated expectations of more time investment, but none demonstrated an increased time investment [21,40,60,61].

Clinician-Patient Relationship

Seventeen descriptive studies reported on the clinician-patient relationship [21,32,35,38-43,45,50,59,62,63,65,71,72]. Patients reported that they were feeling better about their doctors after reading their records [32,39]. They appreciated their doctors' expertise more and experienced a more equal relationship [40,41,43,45,62,64,65,72]. They valued the level of transparency, especially when notes were written respectfully [35,43,50,59]. Respectfully written notes contributed to their feelings of trust [35,71]. As a result, they felt heard and cared for [45]. Three intervention-control studies and 1 pre-post study reported on the professional-client relationship and found no significant differences [44,57,74]. Two other pre-post studies, however, found that the experienced increase of trust in physicians was less than expected, both from a patient and a professional perspective [60,61].

Related to the fear of damaging a therapeutic relationship, some professionals expected that they would report differently if they knew patients could be reading their visit notes. A psychiatrist in one study said: "Sometimes a disbalance occurs, patients 'directing their care' and dictating their doctors how to write their notes" [41]. These psychiatrists also feared that transparency of records could damage the therapeutic relationship, especially when notes revealed subjective impressions. Four pre-post intervention studies investigated if clinicians reported differently about sensitive subjects. Professionals appeared to report less differently than they had expected [21,57,58,61].

Access to Care

An access to care dimension was mentioned in 5 qualitative studies [42,45,49,62,63]. Patients experienced that the PAEHRs gave easy and quick access to health information [42,45,62]. Rapid access was perceived to be advantageous in emergency situations [49]. One study also mentioned that immediacy of secure messaging cultivated a sense of ease of access [63].

Coordination and Continuity

In 2 qualitative studies [40,58] continuity and coordination of care came up. Patients mentioned the benefit of being able to bring their health information along to another care provider and to take care of their own medication when they are out of town.

Differences Among Population Groups

Since former research suggests that different population groups do not profit equally from the use of PAEHRs [19,20,22,23], we searched for differences in our review. Seven studies compared the composition of the studied population with national demographic data. They reported that PAEHR users were more likely to be White and higher educated than nonusers [30,35,36,39,40,44,45]. Four studies investigated experiences of different ethnic and socioeconomic groups [32,45,46,49]. One descriptive study found that women, older patients, and high frequency users found reading notes very important to engaging in their care [45]. Another descriptive study reported that older, lower educated, retired, and unemployed patients, as well as patients with a poor self-reported health and participants in other studies were more willing to share visit notes with others [49]. A third descriptive study found that disadvantaged groups such as the elderly, non-White patients, less educated patients, or patients with poor self-reported health, reported more often than others that use of a PAEHR made them feel better about their doctors [32]. One intervention-control study focused on the importance of PAEHRs to non-White and less educated patients [46]. Both non-White and less educated patients reported more often than White and higher educated patients that the PAEHRs helped them to understand and remember care plans, feel informed, and make decisions concerning their own care. Both non-White patients and less educated patients found reading notes extremely important to engaging in their care.

Discussion

Summary

This review investigates whether and how the use of PAEHRs contributes to patient-centered care, both in general and among specific patient groups. Overall, the articles in this review support the assumption that patient-accessible records contribute to patient-centered care. In all 34 descriptive studies, a positive effect is reported for different dimensions. One descriptive study reported a possible negative effect of PAEHRs on the "therapeutic relationship." Five out of 22 pre-post or intervention-control studies reported significant positive effects related to the dimensions "information," "involvement of patients," or "empowerment." No significant negative effects were reported.

The studies in this review included adults only. Four studies found that, in particular, disadvantaged groups experienced PAEHR-related benefits [32,45,46,49].

Dimensions of Patient-Centered Care

As we expected, the effect on the different "activities" in the Scholl et al model [5] was described most often. Although some effects on "enablers" are reported, only two of the "enablers" are mentioned: (1) access to care [42,45,49,62,63] and (2) coordination/continuity of care [40,58,76]. A complicating factor in the analysis was the varied use of dimensions and their definitions. For instance, whereas Scholl et al [5] distinguished "information," "involvement in care," and "empowerment" as different dimensions, some studies included "involvement" and "knowledge/information" in questionnaires about "empowerment" [5,40,68,71].

Furthermore, we found topics in our review that were not described by Scholl et al [5]. One topic was that patients contributed to patient safety by finding and correcting errors in their records [45,60,62,65,76]. After discussing this topic, we added the subject to "involvement in care," arguing that patients showed their involvement in care by checking their record for errors. In a recent article by Zeh et al [78], however, patient safety was added to the Scholl et al model [5] as a new dimension based on a Delphi study among patients. Patients regarded patient safety as an important dimension of patient-centered care.

Both negative and positive effects were reported for the dimension "patient-clinician relationship." In particular, professionals in mental health care expressed concerns that the transparency of PAEHRs would damage the patient-clinician relationship [38,61]. This is in line with results from other studies. In a recent Norwegian study [79], professionals in mental health care report significantly more often than their colleagues in somatic care that they change their way of writing when using PAEHRs. They also discuss significantly more often than their colleagues in somatic care whether patients should be denied access to their record. Dobscha et al [80] reported that only half of the mental health professionals they queried (107/198) considered sharing mental health Open Notes with patients a good idea, while most of them (174/205) supported the idea in general to share medical notes with patients.

In opposition to professionals, mental health care patients in our review felt that transparency in a PAEHR strengthened the patient-clinician relationship, given that sensitive information was reported in a respectful way [35,38]. The fact that professionals see this differently could be caused by traditional role expectations "in which the patient is viewed as someone to 'protect' and for whom the clinician is responsible" [38]. These role expectations are at odds with the patient-centered care principle of "equal partnership between client and professional" and might cause the reluctance toward the use of transparent PAEHRs.

In line with this assumption, another study emphasizes the importance of a patient-centered attitude by offering specific recommendations for mental health professionals to strengthen the therapeutic alliance in the context of patient-accessible records [35]. These recommendations focus on the "principle" dimensions from the Scholl et al model [5]. The findings in these studies strengthen the assumption in the Scholl et al model that the "activity" dimensions only become visible if the "principles" of patient-centered care, reflected in a patient-centered attitude, have been embraced by professionals.

Differences Among Population Groups

Previous research suggests that disadvantaged groups might profit less from the introduction of PAEHRs than others because they make less use of PAEHRs [19,20,22,23]. In our review, 7 studies reported that users of PAEHRs were more likely to be White and higher educated than nonusers [30,35,36,39,40,44,45], probably due to different access abilities [36]. Surprisingly, 4 other studies found that disadvantaged groups experienced heightened benefits from the use of PAEHRs [32,45,46,49]. An explanation for this benefit could be the value of rereading information that cannot be absorbed all at once. Moreover, Bell et al [32] state that non-White patients are said to distrust White medical professionals, not expecting them to respect their cultural values. Reading transparent records would prove otherwise and might help these patients to trust their doctors more [32]. These findings show that disadvantaged groups benefit from the use of PAEHRs, once they have found their way into the system. This emphasizes the importance in designing and implementing PAEHRs that are easily accessible in order to include disadvantaged groups.

Practical Implications

Our review shows that the use of PAEHRs could enhance patient-centered care, but the effects can be influenced by factors on professional and patient levels. On a professional level, adoption of the principles of patient centered care appears to be crucial for a positive impact of the use of PAEHRs on the patient-clinician relationship. On the patient level, easy access and user-friendliness is important to secure access for all demographics and to facilitate the PAEHR-related benefits that disadvantaged groups might experience.

Strengths and Limitations

One of the strengths of this scoping review is that we included all types of designs and we did not focus on "patient-centered care-specific" search terms. As a result, we created a broad overview on the topic. Subsequently, the analysis was guided by the use of selected dimensions of patient-centered care from Scholl et al [5], which helped us to organize and interpret the information and added strength to the review. On the other hand, the fact that the analysis was conducted in separate dimensions made it more difficult to explore interaction and dependence between the dimensions and to draw conclusions about the impact of PAEHRs on patient-centered care as a whole.

Another strength is the combination of searches from 5 different databases, from both a medical and a social perspective.

A limitation of this review is that, by specifying only "physicians" in our search terms and not "nurses," "nurse practitioners," or nonmedical professionals, we could have missed some articles that were relevant to the subject.

One more limitation of this review is that we included articles in only English and Dutch and no unpublished data or grey literature. For example, no articles from Estonia or Japan could be included, although both countries are very active in eHealth and the government of Estonia has implemented a PAEHR system that is being used for every citizen of the country.

The strength of the conclusions in this review also depends on the quality of the individual studies. Therefore, we conducted a global quality check, where aspects of study design and population were assessed. Although a thorough quality appraisal is not common in scoping reviews, a more detailed quality check could have added strength to the review. The global check indicated that, on average, study results could have been biased because of population selection, as virtually all studies included only native speakers and most of the studies made use of convenience sampling.

Conclusions

This review indicates that PAEHRs bear potential to positively contribute to patient-centered care. However, concerns from professionals about the impact of transparency on the patient-clinician relationship as well as the importance of a patient-centred attitude need to be addressed. Potentially high benefits for disadvantaged groups will be achieved only through easily accessible and user-friendly PAEHRs.

Acknowledgments

This research has partly been made possible by ZonMw, project nr 736300019.

Authors' Contributions

JB, AH, and EV participated in the conception and design of the review. JB, AG, and MG performed the search and elimination of titles/abstracts. All authors contributed to the analysis and interpretation of the data. JB and AH lead the drafting of the manuscript, and EV, AG, and MG contributed to critical revisions of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Completed PRISMA-ScR Checklist.
[PDF File (Adobe PDF File), 499 KB - jmir_v23i1e17655_app1.pdf]

Multimedia Appendix 2
Coding list, used for analysis.
[XLSX File (Microsoft Excel File), 11 KB - jmir_v23i1e17655_app2.xlsx]

Multimedia Appendix 3
Study characteristics, PAEHR functionalities and dimensions of Patient Centered Care.
[XLSX File (Microsoft Excel File), 84 KB - jmir_v23i1e17655_app3.xlsx]

Multimedia Appendix 4
Analysis of outcomes.
[XLSX File (Microsoft Excel File), 33 KB - jmir_v23i1e17655_app4.xlsx]

References

1. Gerties M, Edgman-Levitan S, Daley J, Delbanco TL, editors. Through the patients' eyes: understanding and promoting patient-centred care. San Francisco, CA: Jossey-Bass; Jun 11, 1993.
2. Institute Of Medicine Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; Jul 2001.
3. WHO. People-centred and integrated health services: an overview of the evidence. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/155004/WHO_HIS_SDS_2015.7_eng.pdf?sequence=1 [accessed 2020-03-31]
4. Podlog LW, Brown WJ. Self-determination Theory: A Framework for Enhancing Patient-centered Care. The Journal for Nurse Practitioners 2016 Sep;12(8):e359-e362. [doi: 10.1016/j.nurpra.2016.04.022]
5. Scholl I, Zill JM, Härter M, Dirmaier J. An integrative model of patient-centeredness - a systematic review and concept analysis. PLoS One 2014;9(9):e107828 [FREE Full text] [doi: 10.1371/journal.pone.0107828] [Medline: 25229640]
6. Shaller D. Patient-centered care: What does it take? The Commonwealth Fund. New York: The Commonwealth Fund; 2007 Oct. URL: https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_fund_report_2007_oct_patient_centered_care__what_does_it_take_shaller_patient_centeredcarewhatdoesittake_1067_pdf.pdf
7. Mead N, Bower P. Patient-centred consultations and outcomes in primary care: a review of the literature. Patient Educ Couns 2002 Oct;48(1):51-61. [doi: 10.1016/s0738-3991(02)00099-x] [Medline: 12220750]
8. World Health Organization. People-centered healthcare, a policy framework.: WHO; 2007. URL: https://apps.who.int/iris/bitstream/handle/10665/206971/9789290613176_eng.pdf [accessed 2020-03-31]
9. McMillan SS, Kendall E, Sav A, King MA, Whitty JA, Kelly F, et al. Patient-centered approaches to health care: a systematic review of randomized controlled trials. Med Care Res Rev 2013 Dec;70(6):567-596. [doi: 10.1177/1077558713496318] [Medline: 23894060]
10. Bertakis KD, Azari R. Patient-centered care is associated with decreased health care utilization. J Am Board Fam Med 2011;24(3):229-239 [FREE Full text] [doi: 10.3122/jabfm.2011.03.100170] [Medline: 21551394]
11. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. J Fam Pract 2000 Oct;49(9):796-804. [Medline: 11032203]
12. Bertakis KD, Azari R, Callahan EJ, Helms LJ, Robbins JA. The impact of physician practice style on medical charges. J Fam Pract 1999 Jan;48(1):31-36. [Medline: 9934380]
13. Epstein RM, Franks P, Shields CG, Meldrum SC, Miller KN, Campbell TL, et al. Patient-centered communication and diagnostic testing. Ann Fam Med 2005;3(5):415-421 [FREE Full text] [doi: 10.1370/afm.348] [Medline: 16189057]

14. Dutch Ministry of Justice and Security, Dutch Ministry of Health, Welfare and Sport. Stelselwijziging Jeugd, Hoofdlijnen Wetsvoorstel Jeugdwet. The Hague, the Netherlands: Rijksoverheid; 2013. URL: http://docplayer.nl/19647540-Stelselwijziging-jeugd-hoofdlijnen-wetsvoorstel-jeugdwet.html [accessed 2020-12-31] [WebCite Cache ID http://docplayer.nl/19647540-Stelselwijziging-jeugd-hoofdlijnen-wetsvoorstel-jeugdwet.html]
15. Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc 2014;21(4):737-741 [FREE Full text] [doi: 10.1136/amiajnl-2013-002239] [Medline: 24154835]
16. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
17. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Mar 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
18. Vermeir P, Degroote S, Vandijck D, Van Tiggelen H, Peleman R, Verhaeghe R, et al. The patient perspective on the effects of medical record accessibility: a systematic review. Acta Clin Belg 2017 Jul;72(3):186-194. [doi: 10.1080/17843286.2016.1275375] [Medline: 28056665]
19. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
20. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Oct 08;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
21. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
22. Wouters M, Swinkels I, Lettow BV, Jong JD, Sinnige J, Brabers A. E-health in verschillende snelheden, eHealth-monitor 2018.: Nivel; 2018. URL: https://www.nivel.nl/nl/publicaties/1003379 [accessed 2020-12-31]
23. Lester M, Boateng S, Studeny J, Coustasse A. Personal Health Records: Beneficial or Burdensome for Patients and Healthcare Providers? Perspect Health Inf Manag 2016;13:1h [FREE Full text] [Medline: 27134613]
24. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
25. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
26. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Oct;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
29. Research Square. Jounalguide-Home.: Journalguide.com URL: https://www.journalguide.com/ [accessed 2020-12-31]
30. Aljabri D, Dumitrascu A, Burton MC, White L, Khan M, Xirasagar S, et al. Patient portal adoption and use by hospitalized cancer patients: a retrospective study of its impact on adverse events, utilization, and patient satisfaction. BMC Med Inform Decis Mak 2018 Jul 27;18(1):70 [FREE Full text] [doi: 10.1186/s12911-018-0644-4] [Medline: 30053809]
31. Azizi A, Aboutorabi R, Mazloum-Khorasani Z, Afzal-Aghaea M, Tabesh H, Tara M. Evaluating the Effect of Web-Based Iranian Diabetic Personal Health Record App on Self-Care Status and Clinical Indicators: Randomized Controlled Trial. JMIR Med Inform 2016 Oct 21;4(4):e32 [FREE Full text] [doi: 10.2196/medinform.6433] [Medline: 27769953]
32. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2017 Apr;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
33. Britto MT, Hesse EA, Kamdar OJ, Munafo JK. Parents' perceptions of a patient portal for managing their child's chronic illness. J Pediatr 2013 Jul;163(1):280-1.e1. [doi: 10.1016/j.jpeds.2013.02.041] [Medline: 23541773]
34. Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform 2002 Dec 18;68(1-3):113-127. [doi: 10.1016/s1386-5056(02)00070-9] [Medline: 12467796]
35. Cromer R, Denneson LM, Pisciotta M, Williams H, Woods S, Dobscha SK. Trust in Mental Health Clinicians Among Patients Who Access Clinical Notes Online. PS 2017 May;68(5):520-523. [doi: 10.1176/appi.ps.201600168]

36. Crouch PB, Rose CD, Johnson M, Janson SL. A pilot study to evaluate the magnitude of association of the use of electronic personal health records with patient activation and empowerment in HIV-infected veterans. PeerJ 2015;3:e852 [FREE Full text] [doi: 10.7717/peerj.852] [Medline: 25802815]
37. Kelly MM, Hoonakker PLT, Dean SM. Using an inpatient portal to engage families in pediatric hospital care. J Am Med Inform Assoc 2017 Jan;24(1):153-161 [FREE Full text] [doi: 10.1093/jamia/ocw070] [Medline: 27301746]
38. Denneson LM, Chen JI, Pisciotta M, Tuepker A, Dobscha SK. Patients' Positive and Negative Responses to Reading Mental Health Clinical Notes Online. Psychiatr Serv 2018 May 01;69(5):593-596. [doi: 10.1176/appi.ps.201700353] [Medline: 29493408]
39. Denneson LM, Cromer R, Williams HB, Pisciotta M, Dobscha SK. A Qualitative Analysis of How Online Access to Mental Health Notes Is Changing Clinician Perceptions of Power and the Therapeutic Relationship. J Med Internet Res 2017 Jun 14;19(6):e208 [FREE Full text] [doi: 10.2196/jmir.6915] [Medline: 28615152]
40. Earnest MA, Ross SE, Wittevrongel L, Moore LA, Lin C. Use of a patient-accessible electronic medical record in a practice for congestive heart failure: patient and physician experiences. J Am Med Inform Assoc 2004;11(5):410-417 [FREE Full text] [doi: 10.1197/jamia.M1479] [Medline: 15187074]
41. Elers P, Nelson F. Improving healthcare through digital connection? Findings from a qualitative study about patient portals in New Zealand. Aust J Prim Health 2018 Dec;24(5):404-408. [doi: 10.1071/PY17116] [Medline: 30149829]
42. Esch T, Mejilla R, Anselmo M, Podtschaske B, Delbanco T, Walker J. Engaging patients through open notes: an evaluation using mixed methods. BMJ Open 2016 Jan 29;6(1):e010034 [FREE Full text] [doi: 10.1136/bmjopen-2015-010034] [Medline: 26826154]
43. Fisher B, Bhavnani V, Winfield M. How patients use access to their full health records: a qualitative study of patients in general practice. J R Soc Med 2009 Dec;102(12):539-544 [FREE Full text] [doi: 10.1258/jrsm.2009.090328] [Medline: 19966130]
44. Fossa AJ, Bell SK, DesRoches C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. J Am Med Inform Assoc 2018 Sep 01;25(9):1153-1159 [FREE Full text] [doi: 10.1093/jamia/ocy083] [Medline: 29982659]
45. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What Patients Value About Reading Visit Notes: A Qualitative Inquiry of Patient Experiences With Their Health Information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
46. Gerard M, Chimowitz H, Fossa A, Bourgeois F, Fernandez L, Bell SK. The Importance of Visit Notes on Patient Portals for Engaging Less Educated or Nonwhite Patients: Survey Study. J Med Internet Res 2018 May 24;20(5):e191 [FREE Full text] [doi: 10.2196/jmir.9196] [Medline: 29793900]
47. Gravis G, Protière C, Eisinger F, Boher JM, Tarpin C, Coso D, et al. Full access to medical records does not modify anxiety in cancer patients: results of a randomized study. Cancer 2011 Oct 15;117(20):4796-4804 [FREE Full text] [doi: 10.1002/cncr.26083] [Medline: 21607939]
48. Hanna L, Gill SD, Newstead L, Hawkins M, Osborne RH. Patient perspectives on a personally controlled electronic health record used in regional Australia. Health Inf Manag 2017 Jan;46(1):42-48. [doi: 10.1177/1833358316661063] [Medline: 27486184]
49. Jackson SL, Mejilla R, Darer JD, Oster NV, Ralston JD, Leveille SG, et al. Patients who share transparent visit notes with others: characteristics, risks, and benefits. J Med Internet Res 2014 Dec 12;16(11):e247 [FREE Full text] [doi: 10.2196/jmir.3363] [Medline: 25405911]
50. Kayastha N, Pollak KI, LeBlanc TW. Open Oncology Notes: A Qualitative Study of Oncology Patients' Experiences Reading Their Cancer Care Notes. J Oncol Pract 2018 Apr;14(4):e251-e258. [doi: 10.1200/JOP.2017.028605] [Medline: 29443650]
51. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting Families to Their Health Record and Care Team: The Use, Utility, and Impact of a Client/Family Health Portal at a Children's Rehabilitation Hospital. J Med Internet Res 2017 Apr 06;19(4):e97 [FREE Full text] [doi: 10.2196/jmir.6811] [Medline: 28385680]
52. Mák G, Smith Fowler H, Leaver C, Hagens S, Zelmer J. The Effects of Web-Based Patient Access to Laboratory Results in British Columbia: A Patient Survey on Comprehension and Anxiety. J Med Internet Res 2015 Aug 04;17(8):e191 [FREE Full text] [doi: 10.2196/jmir.4350] [Medline: 26242801]
53. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
54. Nagykaldi Z, Aspy CB, Chou A, Mold JW. Impact of a Wellness Portal on the delivery of patient-centered preventive care. J Am Board Fam Med 2012 Apr;25(2):158-167 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110130] [Medline: 22403196]
55. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to Electronic Health Records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]

56. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
57. Oster NV, Jackson SL, Dhanireddy S, Mejilla R, Ralston JD, Leveille S, et al. Patient Access to Online Visit Notes: Perceptions of Doctors and Patients at an Urban HIV/AIDS Clinic. J Int Assoc Provid AIDS Care 2015 Jul;14(4):306-312 [FREE Full text] [doi: 10.1177/2325957414526783] [Medline: 24729072]
58. Ant Ozok A, Wu H, Garrido M, Pronovost PJ, Gurses AP. Usability and perceived usefulness of Personal Health Records for preventive health care: a case study focusing on patients' and primary care providers' perspectives. Appl Ergon 2014 May;45(3):613-628. [doi: 10.1016/j.apergo.2013.09.005] [Medline: 24119975]
59. Pai HH, Lau F, Barnett J, Jones S. Meeting the health information needs of prostate cancer patients using personal health records. Curr Oncol 2013 Dec;20(6):e561-e569 [FREE Full text] [doi: 10.3747/co.20.1584] [Medline: 24311957]
60. Pell JM, Mancuso M, Limon S, Oman K, Lin C. Patient access to electronic health records during hospitalization. JAMA Intern Med 2015 May;175(5):856-858. [doi: 10.1001/jamainternmed.2015.121] [Medline: 25751393]
61. Petersson L, Erlingsdóttir G. Open Notes in Swedish Psychiatric Care (Part 2): Survey Among Psychiatric Care Professionals. JMIR Ment Health 2018 Jul 21;5(2):e10521 [FREE Full text] [doi: 10.2196/10521] [Medline: 29929946]
62. Rexhepi H, Åhlfeldt RM, Cajander Å, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: A qualitative study. Health Informatics J 2018 Jun 19;24(2):115-124 [FREE Full text] [doi: 10.1177/1460458216658778] [Medline: 27440056]
63. Rief JJ, Hamm ME, Zickmund SL, Nikolajski C, Lesky D, Hess R, et al. Using Health Information Technology to Foster Engagement: Patients' Experiences with an Active Patient Health Record. Health Commun 2017 Mar;32(3):310-319. [doi: 10.1080/10410236.2016.1138378] [Medline: 27223684]
64. Riippa I, Linna M, Rönkkö I. The effect of a patient portal with electronic messaging on patient activation among chronically ill patients: controlled before-and-after study. J Med Internet Res 2014 Dec 19;16(11):e257 [FREE Full text] [doi: 10.2196/jmir.3462] [Medline: 25413368]
65. Ronda MCM, Dijkhorst-Oei L, Vos RC, Rutten GEHM. Diabetes care providers' opinions and working methods after four years of experience with a diabetes patient web portal; a survey among health care providers in general practices and an outpatient clinic. BMC Fam Pract 2018 Jun 21;19(94) [FREE Full text] [doi: 10.1186/s12875-018-0781-y] [Medline: 29929483]
66. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin C. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
67. Shaw E, Howard M, Chan D, Waters H, Kaczorowski J, Price D, et al. Access to Web-Based Personalized Antenatal Health Records for Pregnant Women: A Randomized Controlled Trial. Journal of Obstetrics and Gynaecology Canada 2008 Jan;30(1):38-43. [doi: 10.1016/S1701-2163(16)32711-6] [Medline: 18198066]
68. Tuil WS, Verhaak CM, Braat DDM, de Vries Robbé PF, Kremer JAM. Empowering patients undergoing in vitro fertilization by providing Internet access to medical data. Fertil Steril 2007 Aug;88(2):361-368. [doi: 10.1016/j.fertnstert.2006.11.197] [Medline: 17416366]
69. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. J Am Med Inform Assoc 2014;21(4):657-663 [FREE Full text] [doi: 10.1136/amiajnl-2014-002723] [Medline: 24740865]
70. van den Heuvel SC, Meije D, Regeer EJ, Sinnema H, Riemersma RF, Kupka RW. The user experiences and clinical outcomes of an online personal health record to support self-management of bipolar disorder: A pretest-posttest pilot study. Journal of Affective Disorders 2018 Oct;238:261-268. [doi: 10.1016/j.jad.2018.05.069]
71. van DVR, Drossaert CHC, Taal E, Drossaers-Bakker KW, Vonkeman HE, van DLMAFJ. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
72. Verstraete E, Koehorst AM, van OH. [Does the patient benefit from real-time access to one's electronic record? Evaluation of the patient portal in University Medical Centre Utrecht, the Netherlands]. Ned Tijdschr Geneeskd 2016;160:D325. [Medline: 27299495]
73. Wäckerle A, Blöchlinger-Wegmann B, Burkhardt T, Krähenmann F, Kurmanavicius J, Zimmermann R. Notes on a stick: use and acceptability of woman-held maternity notes. Eur J Obstet Gynecol Reprod Biol 2010 Dec;153(2):156-159. [doi: 10.1016/j.ejogrb.2010.07.019] [Medline: 20674133]
74. Wagner PJ, Dias J, Howard S, Kintziger KW, Hudson MF, Seol Y, et al. Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc 2012 Aug;19(4):626-634 [FREE Full text] [doi: 10.1136/amiajnl-2011-000349] [Medline: 22234404]
75. Wiljer D, Leonard KJ, Urowitz S, Apatu E, Massey C, Quartey NK, et al. The anxious wait: assessing the impact of patient accessible EHRs for breast cancer patients. BMC Med Inform Decis Mak 2010;10:46 [FREE Full text] [doi: 10.1186/1472-6947-10-46] [Medline: 20809950]

76. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2017 May 01;24(e1):e166-e172 [FREE Full text] [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
77. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
78. Zeh S, Christalle E, Hahlweg P, Härter M, Scholl I. Assessing the relevance and implementation of patient-centredness from the patients' perspective in Germany: results of a Delphi study. BMJ Open 2019 Dec 23;9(12):e031741 [FREE Full text] [doi: 10.1136/bmjopen-2019-031741] [Medline: 31874875]
79. Johansen MA, Kummervold PE, Sørensen T, Zanaboni P. Health Professionals' Experience with Patients Accessing Their Electronic Health Records: Results from an Online Survey. Stud Health Technol Inform 2019 Aug 21;264:504-508. [doi: 10.3233/SHTI190273] [Medline: 31437974]
80. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]

Abbreviations

PAEHR: patient accessible health records
PAM-13: 13-question Patient Activation Measurement

Edited by R Kukafka; submitted 01.01.20; peer-reviewed by I Scholl, I Mircheva, M Ameko; comments to author 10.03.20; revised version received 15.06.20; accepted 28.10.20; published 11.01.21.

Please cite as:
Benjamins J, Haveman-Nies A, Gunnink M, Goudkuil A, de Vet E
How the Use of a Patient-Accessible Health Record Contributes to Patient-Centered Care: Scoping Review
J Med Internet Res 2021;23(1):e17655
URL: http://www.jmir.org/2021/1/e17655/
doi:10.2196/17655
PMID:33427683

Review

Young People's Use of Digital Health Technologies in the Global North: Narrative Review

Deborah Lupton[1], PhD

Vitalities Lab, Centre for Social Research in Health and Social Policy Research Centre, UNSW Sydney, Kensington, Australia

Corresponding Author:
Deborah Lupton, PhD
Vitalities Lab
Centre for Social Research in Health and Social Policy Research Centre
UNSW Sydney
Goodsell Building
Kensington, 2052
Australia
Phone: 61 403076791
Email: d.lupton@unsw.edu.au

Abstract

Background: A diverse array of digital technologies are available to children and young people living in the Global North to monitor, manage, and promote their health and well-being.

Objective: This article provides a narrative literature review of the growing number of social research studies published over the past decade that investigate the types of digital technologies used by children and young people in the Global North, in addition to investigating which of these technologies they find most useful or not useful. Key findings as well as major gaps and directions for future research are identified and discussed.

Methods: A comprehensive search of relevant publications listed in Google Scholar was conducted, supported by following citation trails of these publications. The findings are listed under type of digital technology used for health: cross-media, internet, social media, apps and wearable devices, sexual health support and information, and mental health support and information.

Results: Many young people in the Global North are active users of digital health technologies. However, it is notable that they still rely on older technologies, such as websites and search engines, to find information. Apps and platforms that may not have been specifically developed for young people as digital health resources often better suit their needs. Young people appreciate the ready availability of information online, the opportunities to learn more about their bodies and health states, and the opportunities to learn how to improve their health and physical fitness. They enjoy being able to connect with peers, and they find emotional support and relief from distress by using social media platforms, YouTube, and online forums. Young people can find the vast reams of information available to them difficult to navigate. They often look to trusted adults to help them make sense of the information they find online and to provide alternative sources of information and support. Face-to-face interactions with these trusted providers remain important to young people. Risks and harms that young people report from digital health use include becoming overly obsessed with their bodies' shape and size when using self-tracking technologies and comparing their bodies with the social media influencers they follow.

Conclusions: Further details on how young people are using social media platforms and YouTube as health support resources and for peer-to-peer sharing of information, including attention paid to the content of these resources and the role played by young social media influencers and microcelebrities, would contribute important insights to this body of literature. The role played by visual media, such as GIFs (Graphics Interchange Format) and memes, and social media platforms that have recently become very popular with young people (eg, Snapchat and TikTok) in health-related content creation and sharing requires more attention by social researchers seeking to better understand young people's use of digital devices and software for health and fitness.

(J Med Internet Res 2021;23(1):e18286) doi:10.2196/18286

KEYWORDS

digital health; young people; Global North; social research; narrative review

Introduction

Over the past decade, an expanding array of digital technologies have emerged that can be used for promoting or managing young people's health and physical fitness. These devices and media are often referred to as *digital health* technologies. In addition to established digital media, such as online discussion forums and websites about health topics, telemedicine, and exergames, these technologies include new media and devices, such as social media for the discussion of health topics, apps for mobile devices, and wearable devices, including smartwatches or fitness bands that assist with monitoring information about bodily activities and functions [1]. Children and young people are now often encouraged by their parents or teachers to use these devices and software designed to promote their health and fitness [2-5].

In response to these newer digital health technologies, a growing body of social research has developed to investigate how young people living in countries that are often characterized as *the Global North* are engaging with them. The Global North is a term used in the social sciences and humanities to refer to politically stable and wealthy countries or regions that have a western cultural orientation (ie, Western Europe, North America, the United Kingdom, Australia, and New Zealand). As is the case with research, in general, on young people's use of digital technologies [6], far more research has been published—at least in the English language—on their engagement with health-related media and devices in the Global North compared with other regions.

This article presents a narrative review and synthesis of the most recent literature, published from 2010, on young people's use of digital health technologies in the Global North. It includes relevant studies in disciplines such as sociology and media studies that are not often considered in medical or public health research.

The research questions structuring the narrative review were as follows:

1. What digital media and devices are used for health-related purposes by children and young people, aged 5 to 30 years, living in countries in the Global North?
2. Which of these media and devices do they find most useful and helpful?
3. What are the age-related, gendered, socioeconomic, cultural, and geographical dimensions structuring their use or nonuse of digital health?
4. What risks and harms are associated with young people's digital health use?
5. What are the gaps in the social research literature on this topic?

Methods

There are various approaches to narrative literature reviews. The approach adopted here is not the hybrid *systematic narrative review* that is beginning to appear in the body of reviews focusing on medical internet studies. Instead, this narrative review draws from the long-established tradition of literature reviews in social inquiry. The purpose of this type of narrative review is to provide a scholarly summary and interpretation of a body of literature rather than evaluate the validity of a method or a medical intervention, with the intention of integrating and deepening understanding of a particular research topic [7]. The aim is to provide a synthesis of what is known about the chosen topic and identify research gaps [8]. It is particularly useful for synthesizing the scholarly literature in the humanities and social sciences—the disciplines encompassed in this review.

This narrative review method does not attempt to adopt a quantitative approach focused on *rigor*, *avoidance of bias*, and *interrater reliability*, rather it focuses on a qualitative interpretation and synthesis of the state of knowledge in a selected body of literature on a defined topic [7]. It is a different and complementary approach to the systematic review and is an appropriate approach to adopt for a review that is directed at social research [7].

The approach taken for this narrative review is similar, for example, to that undertaken for an assessment of the literature on the role played by gangs in adolescent mental health [9] and the health effects of video gaming [10]. Iterative searches of Google Scholar were conducted to find relevant publications published in English, using the search terms "children," "young people," "teenagers," and "adolescents" in combination with "digital health" and then, more specifically, combined with "health," "digital media," "websites," "telemedicine," "telehealth," "electronic patient records," "social media," "apps," "smartphones," "wearable devices," "exergames," "Facebook," "YouTube," "Instagram," "Pinterest," "Twitter," and "Snapchat." Google Scholar was chosen as the database to search because, unlike other major databases such as Web of Science, PubMed, or Scopus, it is far more inclusive of humanities and social sciences outputs, such as books, book chapters, and journals, which are not included in science- or medical-based databases, while simultaneously including the publications listed in those databases [11].

I read all of the publications found using these searches, including books, book chapters, reports, and journal articles. I then followed the citation trails, identifying and reading outputs that cited the initial identified studies. Only publications appearing since 2010 were included, as I wanted to limit the review to research that considers the most recent digital technologies for health. As an initial search demonstrated, there has been quite a lag in research on young people's use of digital health, with most relevant studies published since 2015. Publications were not included in the review if they did not have a predominantly social research focus or if they did not specifically focus on children or young people's use of digital technologies for health-related purposes (eg, publications about digital health use by parents or health care providers on behalf of young people, or those concerning evaluations of the accuracy of digital health software or devices). Only those publications referring to children or young people living in the Global North were included. Using these criteria, a total of 71 publications were identified for inclusion in this review.

Discussion of the literature review is organized by type of digital technology, beginning with studies presenting a broad overview of digital technology use for health across devices and software, followed by more specific topics: use of cross-media, internet, social media, health apps, and wearable devices. There is also discussion of the literature on sexual health support and information as well as mental health support and information, as these health areas were prominent in the social research on young people's use of digital health. The Discussion section summarizes the key insights and major gaps identified in the literature and suggests directions for future social research studies.

Results

Cross-Media

Only a small number of studies have investigated young people's use of digital technologies for health-related purposes across the range of diverse media and devices currently available to them. One of the earliest is a national representative survey of young people, aged 13 to 18 years, that was conducted in 2014 in the United States [12]. Findings highlighted the continuing importance to young people of face-to-face encounters with family members and health professionals. While the respondents did report searching online for health information, they turned to their parents more often, as well as to doctors and nurses. Furthermore, most respondents reported that they did not frequently use the internet for health information. When they did use the internet for this purpose, they tended to research fitness, diet, or nutrition topics. This survey found that older, wealthier, more active, and healthier young people searched more frequently, as did African American and Hispanic adolescents. It is notable that most respondents did not use social media when searching for health information. A total of 1 in 5 respondents had used YouTube for health topics, which was more than those who used Facebook or Twitter. Among the respondents possessing a smartphone, 29% had downloaded a health app, with fitness or exercise apps being the most common. However, almost half of those who had downloaded a health app hardly ever or never used it. Most respondents had never used a wearable device for health-related purposes.

Another more recent US-based project using qualitative methods found that young people aged 13 to 18 years living in Seattle reported three main uses of digital technology for health: (1) to gather information, such as from medical websites, social media (ie, Facebook and Pinterest), exercise apps, and YouTube for workout information; (2) to share experiences and view others' experiences to gain social support or inspiration; and (3) to engage in self-tracking for health (ie, apps and wearables) [13]. This research found that social media was not a preferred mode to communicate with health care providers for the participants because of privacy and intrusiveness concerns. Online media and apps were considered easy to use, always available, nonjudgmental, and to have options to share information and experiences anonymously. Negative aspects identified by these young people included online health information not always being credible, not being able to afford some health apps and wearable devices, finding it burdensome to enter information into apps, being tempted to always have the phone on and being distracted, technology not being designed for teenagers' use, and peers sharing health information that may be inaccurate or may encourage risk-taking.

Social research in other English-speaking countries in the Global North has indicated high use of the internet, apps, and wearable devices for health-related purposes among young people but also continuing support of long-standing digital platforms and tools. A 2017 study interviewing young Australians, aged 16 to 25 years, found that they valued the convenience, accessibility, detail, and diversity of health information offered by digital media and devices [14]. These young people were active users of older technologies for health, such as search engines and websites, even more than of new technologies, such as social media, apps, and wearable devices. Similar to the respondents in the US survey by Wartella and colleagues [12], these young Australians continued to value personal connections with other people for providing health information and support, including family members and friends as well as medical professionals.

A synthesis of findings across three English-based projects involving both qualitative and quantitative investigations among secondary school students reported on the findings relating to the use of social media, apps, and wearables for health-related purposes [15]. These studies demonstrated that these young people were both critical and vulnerable users of these devices and software. The participants reported a range of positive benefits, particularly in relation to supporting promotion of healthy diet, exercise, and positive body image. However, they also raised concerns about the power of these technologies to shape, influence, and change their health-related behaviors.

Research among secondary school students in South West England in a 2017-2018 study, using a survey as well as qualitative methods, found that the participants were actively using a range of digital devices and software for health-related purposes [16]. Survey data showed that smartphones were the devices most often used to access health information, and half of the respondents were using mobile apps to track their health, diet, or fitness levels. YouTube was the most popular source of health information: 44% of survey respondents reported use of this platform. Fitness was the most popular category of health-related videos watched by participants. Many participants also reported following content by social media influencers about fitness and diet regimens, often organized around the "fitspiration" or "fitspo" hashtags. However, these young people also reported that they considered official websites, such as those offered by government agencies, as the most helpful source for learning about health.

Despite these high levels of digital health use, the participants in these Australian and English studies often noted that it could be difficult to access the accuracy of health information online among the multiple and varied sources available to them and sought help from trusted adults to negotiate this information. However, they demonstrated little notable interest or concern about commercial third parties who might be accessing or exploiting their health data. Participants in the English study did express caution about the people in their lives with whom they knowingly shared their personal health information. Most were willing to share their health data with their parents but to a much lesser extent with health professionals or their friends.

The Internet

Search engines, websites, and online discussion forums were among the earliest forms of online tools for finding and sharing health information. As outlined in the above section, recent research suggests that these resources remain popular for young people in the Global North. Other studies also support these findings. A systematic review of 19, mostly US-based, studies on young people's health-related internet use identified evidence of adolescents seeking preventive health care information, support, and specific information about common health issues, such as sports injuries, flu, chronic diseases, asthma, sexual health, fitness, and infections [17]. The internet served as a confidential source of information that was particularly appreciated by youth facing limited access to health care or managing cultural or religious sensitivities. The value for young people seeking help and support from peers online for mental health problems, living with chronic illnesses, and connecting with young people with the same sexual orientation was a common finding across these studies. Gender and age structured young people's digital health practices. Girls tended to use the internet more often than boys. Older youth preferred online sources of health information, while younger people sought help primarily from their parents, teachers, and other adults. Overall, the young people in these studies perceived their online experiences as positive, but they expressed concern about online privacy and the accuracy of the information they were accessing.

Another systematic review assessed studies of how adolescents search for and evaluate health information online [18]. Four themes were identified across the literature: (1) use of search engines, (2) difficulties in selecting appropriate search strings, (3) barriers to searching, and (4) absence of searching. The review concluded that adolescents are aware of the varying quality of online information about health and exhibit a range of strategies for searching and appraising this information. However, one of the most significant barriers to successfully finding information is the volume of search returns they often encounter, sometimes resulting in them giving up in frustration.

A French survey found that almost half of the young internet users, aged 15 to 30 years, in France searched online for health information. Female participants and those of higher socioeconomic status were more likely to do so [19]. In another more recent French survey involving students aged 18 to 24 years, almost all of the students had searched on the internet for health-related information in the past 12 months, suggesting a growing reliance on the internet among this demographic group in France [20].

Social Media

In the past 5 years, a growing number of social researchers have investigated how young people in the Global North use social media for health-related purposes. Analyses of the content of youth-oriented social media [20-22] and visual media, such as selfies [23,24] and the use of hashtags [25,26] as a way of organizing communities of young people around topics including fitness, sporting activities, health, diet, sexuality, gender identity, and self-care, have received attention in the literature. Researchers have demonstrated that social media influencers, microcelebrities, and other content creators, many of whom are young people themselves, are playing important roles in conveying health information and providing emotional support to young people [14-16,26-29].

The "fitspiration" or "fitspo" hashtags used with images intended to show how physical exercise and training can sculpt bodies, particularly on Instagram, have received particular attention, with researchers showing how young women's bodies are depicted as sexualized, normatively slim, toned, and typically White, while young men are shown as conforming to the muscular or hyper-muscular ideal [30-34]. The relationship between the consumption of digital media and young people's body image has also been explored in several studies, with some evidence to suggest that such images can influence young people's appraisal of, and satisfaction with, their body size, shape, and level of fitness [16,27,35-39]. Researchers have also raised concerns about the promotion of health-detracting practices in social media and YouTube videos, such as self-harm [38] and disordered or overly restricted eating [31,33,39-43].

Goodyear and colleagues used participatory class activities, interviews, focus group discussions, and an online survey with English students, aged 11 to 15 years, to explore their engagement with health-related social media across the range of platforms available to them [44]. Just over half of the survey respondents actively used social media to search for health information, predominantly for physical activity, diet and nutrition, and body image. Almost half (46%) of the respondents said that they had changed their health-related behaviors because of something they had seen on social media; most of them said the change was positive. There were five types of content these participants identified as influential: (1) automatically sourced (ie, pushed to them by the platform's algorithms, including advertising material), (2) suggested or recommended following an initial search, (3) peer generated, (4) likes (ie, positive responses from others regarding their content, suggesting it was good), and (5) reputable content (ie, from official organizations, celebrities, sports figures, big name commercial companies, etc).

A study using focus groups and a survey included adolescent girls living in the US state of West Virginia and queried them about their use of health-related social media [45]. The findings demonstrated the popularity of Snapchat and Instagram with this age group as social media platforms to communicate with friends, with Facebook proving unpopular and viewed as "for older people." In terms of health-related content, these platforms were used to share nutrition information, healthy recipes, weight loss, and fitness posts with friends.

Apps and Wearable Devices

Several systematic reviews have been published of studies investigating various aspects of young people's use of health apps. One focused on studies of adolescents' use of apps for chronic disease management, finding that there was a paucity of evidence-based apps and very few studies evaluating their effectiveness [46]. A systematic review of research on apps used for health promotion among adolescents and students found that most apps were used for therapy or as part of school programs, and only four of them had been specifically designed for adolescents. The review identified that limited research has been conducted on how effective these apps are for promoting health among young people [47].

In-depth research has been able to bring to the surface some of the reasons why young people choose to use health or fitness apps and what aspects of the apps they find most beneficial or helpful. Interviews with American college students who used health and fitness apps found that most of the students had downloaded the app to help with meeting goals, which included supporting an established behavior or adopting a new behavior. They liked apps that were free, easy to use, provided visual and auditory cues, and had game-like rewards, but they did not want to link the apps to social media [48]. Findings from an English study that involved a survey, workshops, and interviews with young people aged 18 to 25 years showed that a calorie-counting app was the most popular and was used by almost half of the participants [49]. Many participants had discontinued using a health or fitness app, however, citing lack of interest, time, or motivation as reasons. Some expressed concern that they were becoming obsessive about counting calories or overexercising, or they struggled with feelings of guilt, anxiety, failure, or disappointment in response to the metrics generated by the app. It was noted by some participants that they felt a need to "reconnect" to their bodies without using digitized and quantified data to generate an understanding of their fitness and health. Similar findings were reported in an interview-based study with young New Zealanders aged 16 to 21 years [50].

Another English-based project involving focus group discussions about fitness apps with adolescent girls who were sports leaders at their school also demonstrated feelings of ambivalence on the part of the girls [51]. These students were mostly positive about using such apps for motivating young people to improve their fitness and health. However, they cautioned that physical education classes should not rely on these kinds of technologies and that the social benefits of exercising with others could be lost.

Far less research has directed attention to young people's use of wearable devices for health and fitness. Studies that have been published suggest that health and fitness apps on smartphones remain more popular than wearable devices among young people. A systematic review of the feasibility and effectiveness of wearable activity trackers among children and young people found that few studies had been conducted. This research suggested that while intervention effects from wearable device use were generally positive, they largely related to short-term rather than long-term effects. Feasibility studies indicated that comfort, design, and feedback features were important [52].

Gender differences were identified in a French survey of students aged 18 to 24 years. The survey found that 35% of the students had used at least one health app, mostly for physical activity and general health monitoring, but only 4% had a wearable device. Female students were more likely to use a health app, while male students were more likely to use a wearable device [20]. In a Belgian survey of adolescent high school students, researchers found that one-quarter of them used nutrition or fitness apps or both. Those who used these types of apps more frequently were more likely to consume healthy beverages and have a lower BMI; however, those who used nutrition apps the most were more likely to have a higher BMI, probably because they were trying to lose weight [53].

Two studies investigated Finnish young people's use of apps and wearable devices for fitness-related purposes. One survey found that half of the respondents said they had these types of apps on their phones, but only 16% reported using them. Only 17% owned wearables and few participants (9%) used them [54]. In the second project, a more specific group of Finnish young people (ie, senior high school athletes) reported a high voluntary use of self-tracking apps and wearable devices to monitor their training and fitness levels, to the point that the use of such technologies was positioned as an important dimension of their identities as accomplished athletes [55].

A small number of projects involved providing young people with fitness trackers to ascertain how they used them. In one such study, Australian adolescents, aged 13 to 14 years, were given a Fitbit Flex wristband and app and were asked to use it for 6 weeks [56]. The participants reported finding it easy for tracking physical activity but not as easy to use for sleep. Barriers to use included its lack of comfort and design, lack of specific feedback about activity levels, and inability to use it in water-based sporting activities. Another Australian study involved providing younger children, aged 7 to 12 years, with an activity tracking device (ie, KidFit) for a period of 4 weeks and included interviews with their parents [57]. Children and parents reported that they found the associated app easy to use; however, the children were frustrated by not being able to receive real-time feedback, and there were difficulties for some with the band feeling uncomfortable and having to remove it for water-based sports activities.

In a US-based research study, young people enrolled in an after-school program were provided with Fitbits for a period of 6 months [58]. Data logs revealed low continuous engagement that declined over time. When the young people did wear their Fitbits, they engaged with their data in reflective ways. The design of the Fitbit, environmental constraints, and motivation were barriers to continuous engagement.

Two projects invited English young people to try Fitbits. One study involved 100 young people aged 13 and 14 years [59,60]. The researchers noted that the participants' motivation to use the wearable devices faded quite quickly, as they became bored with them and disliked the surveillance opportunities the devices afforded their teachers. The young people felt pressured by the demands of the step-counting goals set by the devices. They questioned the value of the metrics generated by their Fitbits and resisted the ways the devices tended to limit physical activity and health status to certain defined measurements, such as step counts.

The second English study adopted a combination of Fitbits and the WhatsApp messenger app to investigate young people's experiences with the devices [16]. A total of 2 girls and 5 boys, aged 16 to 18 years, used the Fitbits for 8 weeks. Several of the participants enjoyed counting their steps, reviewing their metrics instantaneously, competing with others, reaching goals and targets, and receiving positive notification from the devices. For some, the devices generated motivation and short-term behavior changes. Others struggled with knowing how best to respond to the data generated by their Fitbits without becoming too obsessive about their metrics and goals, such as loss of body weight; some simply found the process of remembering to use the devices and to input data or understanding what the metrics were telling them as involving too much labor on their part. As in the other English study, the participants recounted losing the motivation to wear their Fitbits or becoming annoyed with being “nagged” by them to be more physically active.

Sexual Health Support and Information

Health promotion agencies often seek to develop digital resources to support young people’s sexual health. However, these efforts tend to ignore the multiple ways in which a broader ecosystem of digital media and devices are operating in ways that are preferred by young people [61]. These include practices such as consensual sexting, peer-to-peer sharing of information in websites, social media and online discussion forums, and YouTube videos as well as news coverage of sexual health issues on alternative youth-oriented news platforms, such as Buzzfeed, Broadly, and Vice [61-63]. Searching online tends to be the first port of call for young people when they have sexual health concerns [64] or simply want more information about sex [65-67].

Research with LGBTQI (lesbian, gay, bisexual, transgender, queer, and intersex) youth has found that they consider apps that have been specifically designed to support their health and well-being to be “pretty pointless” [68]. In contrast, personal stories are valued by young people looking for information about sexual identity [65]. In particular, the social media platform Tumblr has been an important resource for LGBTQI youth to find connections with others and learn about their sexual identities [69-71]. This opportunity has recently been closed down, however, with Tumblr introducing a policy that it would no longer host “adult content,” thereby effectively censoring much LGBTQI content [69].

Mental Health Support and Information

Many websites, online programs, and apps have been created specifically to support young people’s mental health. A systematic review of studies devoted to the use of mental health apps for children and adolescents showed that very little research has been conducted into their effectiveness [72]. The few studies that have been published failed to demonstrate a significant effect on intended outcomes. Another systematic review focused on research that investigated young people’s online health-seeking behaviors for mental health concerns [73]. Across the included studies, key benefits included anonymity and privacy, immediacy, ease of access, inclusivity, the ability to connect with others and share experiences, and a greater sense of control over the help-seeking journey. Online help-seeking has the potential to meet the needs of those with a preference for self-reliance or act as a gateway to further help-seeking. Barriers to help-seeking included a lack of mental health literacy, concerns about privacy and confidentiality, and uncertainty about the trustworthiness of online resources.

Young people’s mental health and well-being may be supported through digital media that have not been specifically designed for this purpose, and young people themselves often play active roles in generating, curating, and sharing content for these services. For instance, platforms like YouTube, Facebook, Twitter, Tumblr, and Instagram host discussions of topics related to health and well-being, such as mindfulness, stress relief, healthy eating, and physical fitness, as well as provide opportunities for young people to create informal networks to share their experiences of specific mental health conditions and provide information and emotional support to each other [74]. Social research studies have shown that informal online communities developed by young people with peers can promote recovery from conditions such as eating disorders [75], depression [76], suicidal ideation, and self-harm [77].

Online discussion forums have been demonstrated to provide help and support to many young people with mental health conditions or difficulties [78]. A study using focus groups with Australian university students reported that the students appreciated internet-based resources of mental health information and support, particularly for their accessibility and the anonymity and avoidance of stigma these offered [79]. However, they were also concerned about sensitive details about their mental health being accessed by third parties and found the wealth of information distributed across the internet to be difficult to locate and assess for its relevance. A desire for a centralized resource was expressed by these participants.

A Canadian online survey of young people aged 17 to 24 years [80] found that most respondents had used the internet to seek information or help for distressing feelings. Information-based websites were popular, followed by social media sites. Privacy was rated as very important by these Canadian respondents, and they particularly valued information about interventions and treatments. New Zealand–based interviews with a group of participants aged 16 to 21 years noted the importance of mindfulness and meditation apps and trusting relationships with peers on social media for providing relief from distress [50,81].

Discussion

Returning to the research questions addressed by this review, the following trends and gaps were identified across the body of reviewed literature.

What Digital Media and Devices Are Used for Health-Related Purposes by Children and Young People, Aged 5 to 30 Years, Living in Countries in the Global North?

The reviewed research has found that many young people in the Global North are active users of digital health technologies. However, it is notable that they still rely on older technologies, such as websites and search engines, to find information. In many cases, newer media, such as social media platforms, apps, and wearable devices, are less frequently used by young people for health-related purposes. Several studies have demonstrated that young people who have experimented with the use of self-tracking apps and wearable devices do not necessarily continue using them on a long-term basis.

Which of These Media and Devices do They Find Most Useful and Helpful?

Young people in this body of research have reported many benefits from their use of digital technologies for health. They appreciate the ready availability of information on the internet and the opportunities to learn more about their bodies and health states and how to improve their health and physical fitness by accessing websites, discussion forums, and social media and by using apps.

An important finding across the literature is that apps and platforms that may not have been specifically developed for young people as digital health resources (eg, YouTube, Tumblr, and Instagram) often better suit their needs. Young people appreciate being able to connect with peers and find emotional support and relief from distress by using social media platforms, YouTube, and online forums. They can find self-tracking apps and platforms helpful for learning more about their bodies. However, they can become bored, feel overwhelmed by the vast quantities of data generated by these technologies, or find the labor involved in self-tracking as too burdensome for continuing use.

Another key finding across some of the reviewed research is that young people can find the vast reams of information available to them on the internet to be difficult to navigate. They often look to trusted adults, particularly their parents and medical professionals, to help them make sense of the information they find online and to provide alternative sources of information and support. Face-to-face interactions with these trusted providers remain important to young people.

What Are the Aged-Related, Gendered, Socioeconomic, Cultural, and Geographical Dimensions Structuring Young People's Use or Nonuse of Digital Health?

Published studies have mostly been based in the United States and the United Kingdom, but there are also some publications from researchers in Western Europe, Australia, New Zealand, and Canada. Little research thus far has focused on comparing the experiences of young people based on sociodemographic attributes, such as their age, geographical location within a country (eg, in rural or remote regions compared with urban regions), gender, socioeconomic status, and race or ethnicity. Those studies that have made these comparisons have identified that these factors can play an important role in structuring the ways that young people engage with digital health. It is notable that the experiences of preadolescent children have received little attention; most studies published thus far have focused on teenagers and young people aged in their early 20s.

What Risks and Harms Are Associated With Young People's Digital Health Use?

Risks and harms that young people report from digital health use include becoming overly obsessed with their bodies' shape and size using self-tracking technologies and comparing their bodies with the social media influencers they follow. The implications of personal data exploitation for children's rights [82-84] and citizenship [85] have also been noted by media researchers. Major organizations directed at supporting children's rights and well-being, globally, are beginning to draw attention to the opportunities and risks involved in introducing digital technologies into children's lives [86]. Young people often have little choice in engaging with datafication technologies [87]; for example, when they were expected to use digital learning platforms, biometric systems, or self-tracking devices at school, there was little or no option to opt out [2,3,82]. These technologies can limit young people's privacy, agency, and autonomy [88-91].

What Are the Gaps in the Social Research Literature on This Topic?

More comparative research needs to be conducted on how young people from different sociodemographic backgrounds and age groups experience digital health. There is a wide scope for further research that can address these differentiated groups of young people. Only a small number of studies have directly asked young people about their practices or attitudes concerning their health data privacy and security. This research has found that many young people are not overly concerned about these issues, except where it relates to highly sensitive topics, such as sexual or mental health. Young people tend to lack knowledge about the third parties who access their personal health and medical data and what these actors and agencies do with their health information. Given the high prevalence of personal health data hacks, breaches, and leaks [92-94], as well as third-party use of this information across age groups [95-98], what knowledge this age group has about these issues and how they can better learn about and become aware of what happens to their health data require further research attention.

Young people's use of social media and apps has received reasonably high levels of attention by social researchers. However, further details on how young people are using social media platforms and YouTube as health support resources and for peer-to-peer sharing of information, including attention paid to the content of these resources and the role played by young social media influencers and microcelebrities, would contribute important insights to this body of literature.

Finally, the role played by visual media, such as GIFs (Graphics Interchange Format) and memes, and social media platforms that have recently become very popular with young people (eg, Snapchat and TikTok) in health-related content creation and sharing requires more attention by social researchers seeking to better understand young people's use of digital devices and software for health and fitness.

Limitations

This review was limited to research that adopted a social perspective, was published in English, and was conducted with young people living in countries in the Global North. Future reviews could address other bodies of research beyond these parameters to supplement the findings of this review.

Conflicts of Interest

None declared.

References

1. Lupton D. Digital Health: Critical and Cross-Disciplinary Perspectives. London, UK: Routledge; 2017.
2. Gard M, Lupton D. Digital health goes to school: Implications of digitising children's bodies. In: Taylor E, Rooney T, editors. Surveillance Futures: Social and Ethical Implications of New Technologies for Children and Young People. London, UK: Routledge; 2017:36-49.
3. Lupton D. Data assemblages, sentient schools and digitised health and physical education (response to Gard). Sport Educ Soc 2014 Oct 26;20(1):122-132. [doi: 10.1080/13573322.2014.962496]
4. Rich E. Childhood, surveillance and mHealth technologies. In: Taylor E, Rooney T, editors. Surveillance Futures: Social and Ethical Implications of New Technologies for Children and Young People. London, UK: Routledge; 2017:132-146.
5. Rich E. Gender, health and physical activity in the digital age: Between postfeminism and pedagogical possibilities. Sport Educ Soc 2018 Jul 13;23(8):736-747. [doi: 10.1080/13573322.2018.1497593]
6. Livingstone S, Bulger M. A global research agenda for children's rights in the digital age. J Child Media 2014 Sep 24;8(4):317-335. [doi: 10.1080/17482798.2014.961496]
7. Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? Eur J Clin Invest 2018 Jun;48(6):e12931 [FREE Full text] [doi: 10.1111/eci.12931] [Medline: 29578574]
8. Thorne S. On the evolving world of what constitutes qualitative synthesis. Qual Health Res 2019 Jan;29(1):3-6. [doi: 10.1177/1049732318813903] [Medline: 30799767]
9. Macfarlane A. Gangs and adolescent mental health: A narrative review. J Child Adolesc Trauma 2019 Sep;12(3):411-420 [FREE Full text] [doi: 10.1007/s40653-018-0231-y] [Medline: 32318210]
10. Goodman W, McFerran E, Purves R, Redpath I, Beeken R. The untapped potential of the gaming community: Narrative review. JMIR Serious Games 2018 Sep 25;6(3):e10161 [FREE Full text] [doi: 10.2196/10161] [Medline: 30274962]
11. Martín-Martín A, Orduna-Malea E, Thelwall M, Delgado LE. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. J Informetr 2018 Nov;12(4):1160-1177. [doi: 10.1016/j.joi.2018.09.002]
12. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, health and technology: A national survey. Media Commun 2016 Jun 16;4(3):13-23. [doi: 10.17645/mac.v4i3.515]
13. Radovic A, McCarty A, Katzman K, Richardson P. Adolescents' perspectives on using technology for health: Qualitative study. JMIR Pediatr Parent 2018;1(1):e2 [FREE Full text] [doi: 10.2196/pediatrics.8677] [Medline: 30740590]
14. Lupton D. ‘Better understanding about what's going on’: Young Australians’ use of digital technologies for health and fitness. Sport Educ Soc 2018 Dec 06;25(1):1-13. [doi: 10.1080/13573322.2018.1555661]
15. Goodyear V, Armour K. Young people’s perspectives on and experiences of health-related social media, apps, and wearable health devices. Soc Sci 2018 Aug 13;7(8):137. [doi: 10.3390/socsci7080137]
16. Rich E, Lewis S, Lupton D, Miah A, Piwek L. Digital Health Generation? Young People's Use of 'Healthy Lifestyle' Technologies. Bath, UK: University of Bath; 2020. URL: https://static1.squarespace.com/static/59705db2414fb5d8ec26586b/t/5ece2cb7c9d0b7350d143fad/1590570242400/2020DigitalHealthGeneration.pdf [accessed 2020-12-28]
17. Park E, Kwon M. Health-related internet use by children and adolescents: Systematic review. J Med Internet Res 2018 Apr 03;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
18. Freeman JL, Caldwell PH, Bennett PA, Scott KM. How adolescents search for and appraise online health information: A systematic review. J Pediatr 2018 Apr;195:244-255.e1. [doi: 10.1016/j.jpeds.2017.11.031] [Medline: 29398062]
19. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: Results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
20. Montagni I, Cariou T, Feuillet T, Langlois E, Tzourio C. Exploring digital health use and opinions of university students: Field survey study. JMIR Mhealth Uhealth 2018 Mar 15;6(3):e65 [FREE Full text] [doi: 10.2196/mhealth.9131] [Medline: 29549071]
21. Toffoletti K, Thorpe H. Female athletes' self-representation on social media: A feminist analysis of neoliberal marketing strategies in “economies of visibility”. Fem Psychol 2018 Feb 08;28(1):11-31. [doi: 10.1177/0959353517726705]
22. Jenzen O. Trans youth and social media: Moving between counterpublics and the wider web. Gend Place Cult 2017 Nov 13;24(11):1626-1641. [doi: 10.1080/0966369x.2017.1396204]
23. Butkowski C, Dixon T, Weeks K, Smith M. Quantifying the feminine self(ie): Gender display and social media feedback in young women’s Instagram selfies. New Media Soc 2019 Aug 27;22(5):817-837. [doi: 10.1177/1461444819871669]
24. McLean S, Jarman H, Rodgers R. How do “selfies” impact adolescents’ well-being and body confidence? A narrative review. Psychol Res Behav Manag 2019 Jul;12:513-521 [FREE Full text] [doi: 10.2147/prbm.s177834]
25. Ambwani S, Shippe M, Gao Z, Austin S. Is #cleaneating a healthy or harmful dietary strategy? Perceptions of clean eating and associations with disordered eating among young adults. J Eat Disord 2019;7:17 [FREE Full text] [doi: 10.1186/s40337-019-0246-2] [Medline: 31171970]

26. Baker SA, Walsh MJ. ‘Good Morning Fitfam’: Top posts, hashtags and gender display on Instagram. New Media Soc 2018 Jun 10;20(12):4553-4570. [doi: 10.1177/1461444818777514]
27. Goodyear VA, Armour KM, editors. Young People, Social Media and Health. London, UK: Routledge; 2019.
28. Goodyear V. Narrative matters: Young people, social media and body image. Child Adolesc Ment Health 2020 Feb;25(1):48-50 [FREE Full text] [doi: 10.1111/camh.12345] [Medline: 32285640]
29. Baker SA, Rojek C. The Belle Gibson scandal: The rise of lifestyle gurus as micro-celebrities in low-trust societies. J Sociol 2019 May 10;56(3):388-404. [doi: 10.1177/1440783319846188]
30. Carrotte E, Prichard I, Lim M. "Fitspiration" on social media: A content analysis of gendered images. J Med Internet Res 2017 Mar 29;19(3):e95 [FREE Full text] [doi: 10.2196/jmir.6368] [Medline: 28356239]
31. Boepple L, Thompson JK. A content analytic comparison of fitspiration and thinspiration websites. Int J Eat Disord 2016 Jan;49(1):98-101. [doi: 10.1002/eat.22403] [Medline: 25778714]
32. Lupton D. Digital media and body weight, shape, and size: An introduction and review. Fat Stud 2016 Nov 15;6(2):119-134. [doi: 10.1080/21604851.2017.1243392]
33. Tiggemann M, Zaccardo M. 'Strong is the new skinny': A content analysis of #fitspiration images on Instagram. J Health Psychol 2018 Jul;23(8):1003-1011. [doi: 10.1177/1359105316639436] [Medline: 27611630]
34. Camacho-Miñano MJ, MacIsaac S, Rich E. Postfeminist biopedagogies of Instagram: Young women learning about bodies, health and fitness. Sport Educ Soc 2019 May 22;24(6):651-664. [doi: 10.1080/13573322.2019.1613975]
35. Åberg E, Koivula A, Kukkonen I. A feminine burden of perfection? Appearance-related pressures on social networking sites. Telematics Inform 2020 Mar;46:101319. [doi: 10.1016/j.tele.2019.101319]
36. Burnette CB, Kwitowski MA, Mazzeo SE. "I don't need people to tell me I'm pretty on social media:" A qualitative study of social media and body image in early adolescent girls. Body Image 2017 Dec;23:114-125. [doi: 10.1016/j.bodyim.2017.09.001] [Medline: 28965052]
37. Goodyear V, Quennerstedt M. #Gymlad - Young boys learning processes and health-related social media. Qual Res Sport Exerc Health 2019 Oct 07;12(1):18-33. [doi: 10.1080/2159676x.2019.1673470]
38. Fardouly J, Vartanian LR. Social media and body image concerns: Current research and future directions. Curr Opin Psychol 2016 Jun;9:1-5. [doi: 10.1016/j.copsyc.2015.09.005]
39. Holland G, Tiggemann M. A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 2016 Jun;17:100-110. [doi: 10.1016/j.bodyim.2016.02.008] [Medline: 26995158]
40. Holland G, Tiggemann M. "Strong beats skinny every time": Disordered eating and compulsive exercise in women who post fitspiration on Instagram. Int J Eat Disord 2017 Jan;50(1):76-79. [doi: 10.1002/eat.22559] [Medline: 27302867]
41. Ging D, Garvey S. ‘Written in these scars are the stories I can’t explain’: A content analysis of pro-ana and thinspiration image sharing on Instagram. New Media Soc 2017 Jan 20;20(3):1181-1200. [doi: 10.1177/1461444816687288]
42. Marcus S. Thinspiration vs thicksperation: Comparing pro-anorexic and fat acceptance image posts on a photo-sharing site. Cyberpsychology 2016 Jul 01;10(2):1 [FREE Full text] [doi: 10.5817/cp2016-2-5]
43. De Choudhury M. Anorexia on Tumblr: A characterization study. In: Proceedings of the 5th International Conference on Digital Health. 2015 Presented at: 5th International Conference on Digital Health; May 18-20, 2015; Florence, Italy p. 43-50. [doi: 10.1145/2750511.2750515]
44. Goodyear VA, Armour KM, Wood H. Young people and their engagement with health-related social media: New perspectives. Sport Educ Soc 2019;24(7):673-688 [FREE Full text] [doi: 10.1080/13573322.2017.1423464] [Medline: 31814804]
45. Leary M, Clegg E, Santella M, Murray P, Downs J, Olfert M. Consumption of health-related content on social media among adolescent girls: Mixed-methods pilot study. JMIR Form Res 2019 Mar 01;3(1):e11404 [FREE Full text] [doi: 10.2196/11404] [Medline: 30821696]
46. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: A systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
47. Dute J, Bemelmans E, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: A review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 05;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
48. Gowin M, Cheney M, Gwin S, Franklin Wann T. Health and fitness app use in college students: A qualitative study. Am J Health Educ 2015 Jul 06;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
49. Honary M, Bell B, Clinch S, Wild S, McNaney R. Understanding the role of healthy eating and fitness mobile apps in the formation of maladaptive eating and exercise behaviors in young people. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e14239 [FREE Full text] [doi: 10.2196/14239] [Medline: 31215514]
50. Trnka SH. Digital care: Agency and temporality in young people’s use of health apps. Engag Sci Technol Soc 2016 Dec 04;2:248-265 [FREE Full text] [doi: 10.17351/ests2016.119]
51. Depper A, Howe PD. Are we fit yet? English adolescent girls’ experiences of health and fitness apps. Health Sociol Rev 2016 Jul 04;26(1):98-112. [doi: 10.1080/14461242.2016.1196599]
52. Ridgers D, McNarry A, Mackintosh A. Feasibility and effectiveness of using wearable activity trackers in youth: A systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]

53. De Cock N, Vangeel J, Lachat C, Beullens K, Vervoort L, Goossens L, et al. Use of fitness and nutrition apps: Associations with body mass index, snacking, and drinking habits in adolescents. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e58 [FREE Full text] [doi: 10.2196/mhealth.6005] [Medline: 28442455]
54. Ng K, Tynjälä J, Kokko S. Ownership and use of commercial physical activity trackers among Finnish adolescents: Cross-sectional study. JMIR Mhealth Uhealth 2017 May 04;5(5):e61 [FREE Full text] [doi: 10.2196/mhealth.6940] [Medline: 28473304]
55. Ng K, Ryba T. The quantified athlete: Associations of wearables for high school athletes. Adv Hum Comput Interact 2018 Oct 01;2018:1-8 [FREE Full text] [doi: 10.1155/2018/6317524]
56. Ridgers D, Timperio A, Brown H, Ball K, Macfarlane S, Lai K, et al. Wearable activity tracker use among Australian adolescents: Usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
57. Mackintosh K, Chappel S, Salmon J, Timperio A, Ball K, Brown H, et al. Parental perspectives of a wearable activity tracker for children younger than 13 years: Acceptability and usability study. JMIR Mhealth Uhealth 2019 Nov 04;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]
58. Schaefer SE, Ching CC, Breen H, German JB. Wearing, thinking, and moving: Testing the feasibility of fitness tracking with urban youth. Am J Health Educ 2016 Jan 08;47(1):8-16. [doi: 10.1080/19325037.2015.1111174]
59. Goodyear VA, Kerner C, Quennerstedt M. Young people’s uses of wearable healthy lifestyle technologies: Surveillance, self-surveillance and resistance. Sport Educ Soc 2017 Sep 22;24(3):212-225. [doi: 10.1080/13573322.2017.1375907]
60. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: A self-determination perspective on Fitbits with adolescents. Am J Health Educ 2017 Jul 28;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
61. Albury K. ‘Recognition of competition’ versus will to app: Rethinking digital engagement in Australian youth sexual health promotion policy and practice. Media Int Aust 2019 Jun 04;171(1):38-50. [doi: 10.1177/1329878x19850224]
62. McKee A. Learning from commercial entertainment producers in order to create entertainment sex education. Sex Educ 2016 Sep 11;17(1):26-40. [doi: 10.1080/14681811.2016.1228528]
63. Mowat H, Dobson AS, McDonald K, Fisher J, Kirkman M. “For myself and others like me”: Women’s contributions to vulva-positive social media. Fem Media Stud 2018 Nov 19;20(1):35-52. [doi: 10.1080/14680777.2018.1546209]
64. Freeman G, Smith LW, McNulty A, Donovan B. Sexual health and students: The pathways travelled by those with sexual health concerns. Sex Health 2018;15(1):76-78. [doi: 10.1071/sh17083]
65. Starling MS, Cheshire C, Deardorff J, Nuru-Jeter A. Late adolescent user experiences with online sexual health resources: A qualitative study. Am J Sex Educ 2018 Oct 08;13(4):487-503. [doi: 10.1080/15546128.2018.1505572]
66. Dolcini MM, Warren J, Towner SL, Catania JA, Harper GW. Information age: Do urban African American youth find sexual health information online? Sex Res Social Policy 2015 Jun 01;12(2):110-114 [FREE Full text] [doi: 10.1007/s13178-014-0174-5] [Medline: 26052264]
67. Döring N, Daneback K, Shaughnessy K, Grov C, Byers ES. Online sexual activity experiences among college students: A four-country comparison. Arch Sex Behav 2017 Aug;46(6):1641-1652. [doi: 10.1007/s10508-015-0656-4] [Medline: 26659580]
68. Byron P. ‘Apps are cool but generally pretty pointless’: LGBTIQ+ young people’s mental health app ambivalence. Media Int Aust 2019 Apr 26;171(1):51-65. [doi: 10.1177/1329878x19844034]
69. Byron P, Robards B, Hanckel B, Vivienne S, Churchill B. “Hey, I’m having these experiences”: Tumblr use and young people’s queer (dis)connections. Int J Commun 2019;13:2239-2259 [FREE Full text]
70. Byron P. ‘How could you write your name below that?’ The queer life and death of Tumblr. Porn Stud 2019 Jun 21;6(3):336-349. [doi: 10.1080/23268743.2019.1613925]
71. Cavalcante A. Tumbling into queer utopias and vortexes: Experiences of LGBTQ social media users on Tumblr. J Homosex 2019;66(12):1715-1735. [doi: 10.1080/00918369.2018.1511131] [Medline: 30235077]
72. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: A systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
73. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: Systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
74. Hendry NA, Robards B, Stanford S. Beyond social media panics for 'at risk' youth in mental health practice. In: Stanford S, Sharland E, Rovinelli Heller N, Warner J, editors. Beyond the Risk Paradigm in Mental Health Policy and Practice. London, UK: Red Globe Press; 2017:135-154.
75. LaMarre A, Rice C. Hashtag recovery: #Eating disorder recovery on Instagram. Soc Sci 2017 Jun 29;6(3):68 [FREE Full text] [doi: 10.3390/socsci6030068]
76. Andalibi N, Ozturk P, Forte A. Sensitive self-disclosures, responses and social support on Instagram: The case of #depression. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York, NY: Association for Computing Machinery; 2017 Presented at: 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; February 25-March 1, 2017; Portland, OR p. 1485-1500 URL: https://dl.acm.org/doi/pdf/10.1145/2998181.2998243 [doi: 10.1145/2998181.2998243]

77. McCosker A, Hartup M. Turning to Online Peer Forums for Suicide and Self-Harm Support: “It Does Help Having You Guys”. Melbourne, Australia: beyondblue and Swinburne Social Innovation Research Institute; 2017. URL: https://www.beyondblue.org.au/docs/default-source/research-project-files/mccosker-hartup_turning-to-online-forums_beyondblue_suicide-and-self-harm_2017.pdf?sfvrsn=e600adea_2 [accessed 2020-12-28]
78. Webb M, Burns J, Collin P. Providing online support for young people with mental health difficulties: Challenges and opportunities explored. Early Interv Psychiatry 2008 May;2(2):108-113. [doi: 10.1111/j.1751-7893.2008.00066.x] [Medline: 21352141]
79. Chan K, Farrer M, Gulliver A, Bennett K, Griffiths M. University students' views on the perceived benefits and drawbacks of seeking help for mental health problems on the internet: A qualitative study. JMIR Hum Factors 2016 Jan 19;3(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.4765] [Medline: 27026140]
80. Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: A survey of youths' web-based resource preferences in Canada. J Med Internet Res 2014 Dec 17;16(12):e293 [FREE Full text] [doi: 10.2196/jmir.3526] [Medline: 25519847]
81. Gibson K, Trnka S. Young people's priorities for support on social media: “It takes trust to talk about these issues”. Comput Human Behav 2020 Jan;102:238-247. [doi: 10.1016/j.chb.2019.08.030]
82. Lupton D, Williamson B. The datafied child: The dataveillance of children and implications for their rights. New Media Soc 2017 Jan 23;19(5):780-794. [doi: 10.1177/1461444816686328]
83. Livingstone S, Third A. Children and young people’s rights in the digital age: An emerging agenda. New Media Soc 2017 May 10;19(5):657-670. [doi: 10.1177/1461444816686318]
84. Third A, Bellerose D, Dawkins U, Keltie E, Pihl K. Children’s Rights in the Digital Age: A Download From Children Around the World. 2nd edition. Melbourne, Australia: Young and Well Cooperative Research Centre; 2014 Oct. URL: http://www.aeema.net/WordPress/wp-content/uploads/2014/10/Childrens-Rights-in-the-Digital-Age.pdf [accessed 2020-12-28]
85. Mascheroni G. Researching datafied children as data citizens. J Child Media 2018 Sep 12;12(4):517-523. [doi: 10.1080/17482798.2018.1521677]
86. Campo S, Raymond N. Displaced Children and Emerging Technologies: Save the Children’s Opportunities for Investment and Impact. London, UK: Save the Children; 2019. URL: https://resourcecentre.savethechildren.net/node/15382/pdf/stc_tech_innovation_study_v7_digital.pdf [accessed 2020-12-28]
87. Barassi V. Datafied citizens in the age of coerced digital participation. Sociol Res Online 2019 Jun 28;24(3):414-429. [doi: 10.1177/1360780419857734]
88. Barron CM. 'I had no credit to ring you back': Children’s strategies of negotiation and resistance to parental surveillance via mobile phones. Surveill Soc 2014 Jun 17;12(3):401-413. [doi: 10.24908/ss.v12i3.4966]
89. Gabriels K. ‘I keep a close watch on this child of mine’: A moral critique of other-tracking apps. Ethics Inf Technol 2016 Jun 21;18(3):175-184. [doi: 10.1007/s10676-016-9405-1]
90. Bettany SM, Kerrane B. The socio-materiality of parental style. Eur J Mark 2016 Nov 14;50(11):2041-2066. [doi: 10.1108/ejm-07-2015-0437]
91. Livingstone S, Stoilova M, Nandagiri R. Children's Data and Privacy Online: Growing Up in a Digital Age: An Evidence Review. London, UK: LSE Media and Communications; 2018 Dec. URL: https://www.lse.ac.uk/media-and-communications/assets/documents/research/projects/childrens-privacy-online/Evidence-review-final.pdf [accessed 2020-12-28]
92. Argaw S, Bempong N, Eshaya-Chauvin B, Flahault A. The state of research on cyberattacks against hospitals and available best practice recommendations: A scoping review. BMC Med Inform Decis Mak 2019 Jan 11;19(1):10 [FREE Full text] [doi: 10.1186/s12911-018-0724-5] [Medline: 30634962]
93. Millard WB. Where bits and bytes meet flesh and blood. Ann Emerg Med 2017 Sep;70(3):A17-A21. [doi: 10.1016/j.annemergmed.2017.07.008]
94. Dyer O. Hackers demand ransom to release encrypted US medical records. BMJ 2016 Apr 01;353:i1876. [doi: 10.1136/bmj.i1876] [Medline: 27037196]
95. Ebeling M. Healthcare and Big Data: Digital Specters and Phantom Objects. New York, NY: Palgrave Macmillan US; 2016.
96. Lupton D. The commodification of patient opinion: The digital patient experience economy in the age of big data. Sociol Health Illn 2014 Jul;36(6):856-869. [doi: 10.1111/1467-9566.12109] [Medline: 24443847]
97. Sharon T. When digital health meets digital capitalism, how many common goods are at stake? Big Data Soc 2018 Dec 20;5(2):1-12 [FREE Full text] [doi: 10.1177/2053951718819032]
98. Grundy Q, Chiu K, Bero L. Commercialization of user data by developers of medicines-related apps: A content analysis. J Gen Intern Med 2019 Dec;34(12):2833-2841 [FREE Full text] [doi: 10.1007/s11606-019-05214-0] [Medline: 31529374]

Abbreviations

GIF: Graphics Interchange Format
LGBTQI: lesbian, gay, bisexual, transgender, queer, and intersex

Edited by G Eysenbach; submitted 17.02.20; peer-reviewed by V Goodyear, L McCann, A Hoskins; comments to author 17.07.20; revised version received 02.08.20; accepted 28.10.20; published 11.01.21.

Please cite as:
Lupton D
Young People’s Use of Digital Health Technologies in the Global North: Narrative Review
J Med Internet Res 2021;23(1):e18286
URL: http://www.jmir.org/2021/1/e18286/
doi:10.2196/18286
PMID:33427684

Review

The Role of Health Technologies in Multicomponent Primary Care Interventions: Systematic Review

Geronimo Jimenez[1,2], MA; David Matchar[3], MD; Choon Huat Gerald Koh[4], MD, PhD; Rianne van der Kleij[2], PhD; Niels H Chavannes[2], MD, PhD; Josip Car[1], MD, PhD

[1]Centre for Population Health Sciences (CePHaS), Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

[2]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[3]Programme in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

[4]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Corresponding Author:
Geronimo Jimenez, MA
Centre for Population Health Sciences (CePHaS)
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road, Level 18
Clinical Sciences Building, Novena Campus
Singapore, 308232
Phone: 65 6904 7027
Email: geronimo.jimenez@gmail.com

Abstract

Background: Several countries around the world have implemented multicomponent interventions to enhance primary care, as a way of strengthening their health systems to cope with an aging chronically ill population and rising costs. Some of these efforts have included technology-based enhancements as one of the features to support the overall intervention, but their details and impacts have not been explored.

Objective: This study aimed to identify the role of digital/health technologies within wider multifeature interventions that are aimed at enhancing primary care, and to describe their aims and stakeholders, types of technologies used, and potential impacts.

Methods: A systematic review was performed following Cochrane guidelines. An electronic search, conducted on May 30, 2019, was supplemented with manual and grey literature searches in December 2019, to identify multicomponent interventions that included at least one technology-based enhancement. After title/abstract and full text screening, selected articles were assessed for quality based on their study design. A descriptive narrative synthesis was used for analysis and presentation of the results.

Results: Of 37 articles, 14 (38%) described the inclusion of a technology-based innovation as part of their multicomponent interventions to enhance primary care. The most commonly identified technologies were the use of electronic health records, data monitoring technologies, and online portals with messaging platforms. The most common aim of these technologies was to improve continuity of care and comprehensiveness, which resulted in increased patient satisfaction, increased primary care visits compared to specialist visits, and the provision of more health prevention education and improved prescribing practices. Technologies seem also to increase costs and utilization for some parameters, such as increased consultation costs and increased number of drugs prescribed.

Conclusions: Technologies and digital health have not played a major role within comprehensive innovation efforts aimed at enhancing primary care, reflecting that these technologies have not yet reached maturity or wider acceptance as a means for improving primary care. Stronger policy and financial support, and advocacy of key stakeholders are needed to encourage the introduction of efficient technological innovations, which are backed by evidence-based research, so that digital technologies can fulfill the promise of supporting strong sustainable primary care.

(J Med Internet Res 2021;23(1):e20195) doi:10.2196/20195

KEYWORDS

digital health; health system improvements; health technologies; primary care; systematic review

Introduction

Primary care is often considered a cornerstone of health care systems. Health systems with strong primary health care produce better and more equitable health outcomes, are more efficient, and can achieve higher user satisfaction in comparison to health systems with only a weak primary care orientation [1,2]. Changing demographics, an increasingly aging population, and the increased burden of noncommunicable diseases have been identified as new challenges for health systems worldwide [3-5], and strengthening primary care has been proposed as one solution to address these challenges.

Many countries have implemented a wide array of innovations to enhance primary care, ranging from policy initiatives, such as capitated reimbursement, to ground level improvements, such as improving access to primary care practices and enhancing the roles of nurses to provide comprehensive primary care services [6-8]. As in other fields, such as finance, retail, and agriculture, an increasingly important domain for innovation involves the incorporation of technology. Technologies are having an impact on health service delivery and health system administration, and they promise to provide solutions for improving primary care [9,10].

There have been many studies emphasizing individual digital technologies for improving specific aspects of health care and primary care. Some of these include digital health assistants to help with administrative tasks, medical chatbots to engage patients more frequently, and the use of electronic health records and telemedicine, among others [9-11]. However, no studies have explored the role of technologies within multicomponent efforts to enhance primary care, that is, whether within initiatives comprised of several features aimed at enhancing primary care, there was a technology element being introduced, and if yes, what it was.

We aimed to systematically explore the role that health/digital technologies have played in multicomponent efforts designed to improve primary care by identifying (1) the types of technologies implemented, (2) the functional objective of the technology, (3) the relevant stakeholders, and (4) whether they have an impact on enhancing the defining features of primary care (ie, first contact, comprehensiveness, coordination, and continuity) [12], denoted here as the “4Cs.” We explored the overall outcomes of the multicomponent interventions in which technology is one component to attempt to discern the specific contribution of the technologies within these efforts. Textbox 1 provides useful definitions for concepts and terms that will be used throughout the article.

Textbox 1. Useful definitions.

Multicomponent interventions/innovation environments: programs or strategies composed of several innovations/features to enhance primary care.

Innovation features: individual innovation elements included in multicomponent interventions.

Health technologies: application of scientific knowledge to solve health care–related problems, including its corresponding machinery and equipment (includes information technology, digital health, eHealth, mHealth, etc).

4Cs: the primary care core functions (first contact, comprehensiveness, coordination, and continuity).

Quadruple aim outcomes: the four types of outcomes to measure successful health system improvements (population health outcomes, health care utilization and cost outcomes, patient satisfaction, and provider satisfaction).

Methods

A systematic review was designed and performed following Cochrane guidelines for conducting systematic reviews [13]. The detailed methods for this review are described in an article that explored multicomponent interventions aimed at enhancing primary care, which identified 18 innovation strategies and provided a broader picture of the many innovation features used internationally to improve various aspects of primary care simultaneously [14]. A summary is provided below.

An electronic database search was performed in order to identify (1) multicomponent interventions or “innovation environments” aimed at enhancing primary care (with at least three innovation features); (2) factors influencing at least one of the primary care core functions (4Cs), and (3) studies reporting on any of the four basic types of outcomes of a successful health system (the so-called “quadruple aim” outcomes of population health, health care costs and utilization, patient satisfaction, and provider satisfaction) [15] and providing numerical values for at least five outcome measures. In a previous scan of the literature, we identified many specific interventions aimed at enhancing a particular aspect of primary care services (eg, *the paper stamp checklist tool enhances asthma guideline knowledge and implementation by primary care physicians*), and based on this, we determined that consideration of studies describing interventions with at least three distinct innovation features and measuring at least five outcome measures could ensure that the interventions were indeed “multicomponent.”

A search strategy was developed, and it focused on the following three main sets of terms: (1) primary care–related terms; (2) innovation/reform/enhancement-related terms; and (3) study design filters (Multimedia Appendix 1). The electronic database search was performed in Ovid/MEDLINE on May 30, 2019, and it was supplemented by manual searches through the references of the included studies and by a grey literature search (ie, search through materials and documents produced by organizations outside of the traditional commercial or academic publishing and distribution channels, such as government and industry documents) in OpenGrey [16], using “primary care” and “innovation,” on December 12, 2019. From the studies fulfilling these criteria, we selected those that had technology-based enhancements as part of the elements in their multicomponent interventions.

We defined health technologies, using definitions from two World Health Organization reports, as the “application of scientific knowledge for practical purposes, including its corresponding machinery and equipment, to solve health care–related problems and improve quality of life” [17] and encompassing digital health technologies (the overarching term to include eHealth and mHealth, eg, telemedicine, electronic health records [EHRs], and wearable sensors) and their corresponding medical and assistive devices [9].

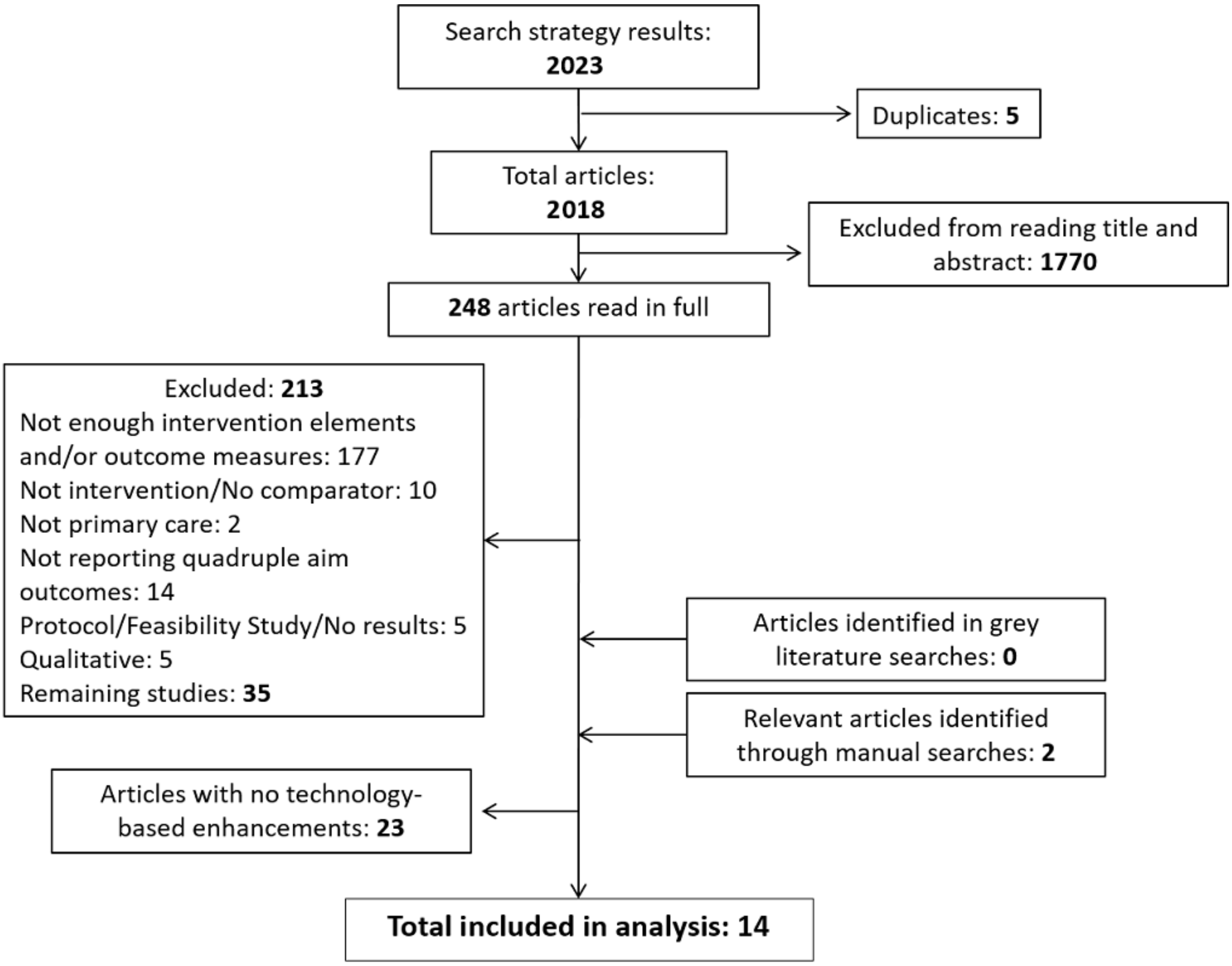
Quality evaluation of the included studies was based on study design, using the National Institutes of Health–National Health, Lung and Blood Institute’s “Study Quality Assessment Tools” [18], a comprehensive suite of study evaluation tools, which has been used in a variety of systematic reviews [19-21]. Data extraction was performed using a predefined data extraction form for study characteristics and general information (author/year, setting/country, policy influence, study design and quality, and patient population involved), primary care intervention elements, and quadruple aim outcomes, including reported magnitudes for each outcome measure. A narrative descriptive approach was utilized to identify and report the types and specific details of the implemented technologies, the involved stakeholders, whether and which 4Cs were arguably supported, and the outcomes influenced by the corresponding technology.

Results

Search Results

After the electronic search, subjecting the articles to the inclusion/exclusion criteria and manual reference and grey literature searches resulted in 37 articles fulfilling the requirements for multicomponent interventions as described above. From these, 14 studies had technology-based enhancements and were included in the analysis (Figure 1).

Figure 1. PRISMA diagram describing the study selection process. Figure extended from Jimenez et al [14].



Study Characteristics

Articles were published between 2008 and 2017, and half of them were published since 2016. Most described studies performed in the United States (9/14, 64%). Additionally, four were from Europe (two from Germany and two from Spain) and one was from Argentina. Eight articles mentioned policies influencing the implementation of the innovation programs as broader country, regional, or organizational efforts to enhance primary care (Table 1) [22-35].

Table 1. Study characteristics organized by study type (N=14) (adapted from Jimenez et al [14]).

Study type	Author (Year)	Program name	Setting/context	Policy/ government program influencing innovation	Study design (quality evaluation rating)[a]	Patient population (if any)	Innovation elements included in the full intervention[b]	Types of outcomes studied
Controlled intervention study	Coderch et al (2016) [22]	__[c]	Integrated health care organization in the region of Girona, Spain in 2011 (128,000 residents)	Catalonia's 2011-2015 health plan; creation of the Program for Chronic Condition Prevention and Care	Controlled, pragmatic, randomized clinical trial, with three arms: one blind control and two open intervention groups (fair)	Complex chronic patients who account for 5% of the highest risk of highest health costs each year	- Accountability mechanisms - Care plan development - Improved access - Improved specialty care access - Enhanced coordination/information exchange efforts - Provider education or training - Technology enhancements	HC[d] costs and utilization
Controlled intervention study	Prestes et al (2017) [23]	DIAPREM study	Primary care units of La Matanza County, Argentina	—	Random selection of 30 PC[e] providers and 30 nurses from 40 PC units (fair)	T2DM[f] patients	- Efforts to improve performance monitoring - Enhanced continuity/transition-based efforts - Provider education or training - Technology enhancements	Population health HC costs and utilization
Controlled intervention study	Ruescas-Escolano et al (2014) [24]	PROP-RESE Trial	Multicentric, PC study (15 health centers), participating in the Cardiometabolic Valencian Study	—	Open randomized clinical trial with 1-year follow-up (good)	Patients with ischemic heart disease	- Efforts to improve performance monitoring - Improved patient self-management - Provider education or training - Others - Team-based care - Technology enhancements	Population health
Observational cohort or cross-sectional study	Dale et al (2016) [25]	Comprehensive Primary Care (CPC) Initiative	A large and diverse set of practices in seven Center for Medicare and Medicaid Services (CMS) regions (four states and three metropolitan regions in the United States)	Launching of the Centers for Medicare and Medicaid Services' Comprehensive Primary Care Initiative, in October 2012	Pre-post design with a comparison site (fair, classified as retrospective cohort for quality evaluation)	Medicare fee-for-service beneficiaries	- Care development plan - Case management - Improved access - Improved patient self-management - Payment-based enhancements - Social or community services engagement - Technology enhancements	HC costs and utilization Patient satisfaction

Study type	Author (Year)	Program name	Setting/context	Policy/ government program influencing innovation	Study design (quality evaluation rating)[a]	Patient population (if any)	Innovation elements included in the full intervention[b]	Types of outcomes studied
Observational cohort or cross-sectional study	Goff et al (2017) [26]	Buena Salud	Program implemented at Brightwood Health Center (BHC) in MA, an urban community health center with a largely Hispanic population (88%) insured primarily by either Medicaid (59%) or Medicare (28%)	—	Controlled before-and-after study (fair)	T2DM patients enrolled in the Buena Salud program	- Accountability mechanisms - Case management - Improved access - Improved patient self-management - Improved specialty care access - Social or community services engagement - Team-based care - Technology enhancements	Population health
Observational cohort or cross-sectional study	Maeng et al (2013) [27]	Proven-Health Navigator	36 Geisinger-owned PC practices, as well as seven contracted PC practices in GHP's[g] provider network. Geisinger's regional health care system is a provider to central, south-central, and northeastern Pennsylvania and southern New Jersey	PCMH[h] transformation in primary care	Survey of patients in "PHN[i] sites." A comparable survey of patients from non-PHN sites was conducted for comparison (fair)	General patient population of PC practices enrolled in the PHN program	- Case management - Efforts to improve performance monitoring - Enhanced service capacity - Improved access - Improved patient self-management - Payment-based enhancements - Social or community services engagement - Team-based care - Technology enhancements	Patient satisfaction
Observational cohort or cross-sectional study	Maeng et al (2012) [28]	Proven-Health Navigator	36 Geisinger-owned PC practices, as well as seven contracted PC practices in GHP's provider network. Geisinger's regional health care system is a provider to central, south-central, and northeastern Pennsylvania and southern New Jersey	PCMH transformation in primary care	Multivariate logistic regression models with controls (members not in the program) (fair, classified as retrospective cohort for quality evaluation)	General patient population of PC practices enrolled in the PHN program	- Case management - Efforts to improve performance monitoring - Enhanced service capacity - Improved access - Improved patient self-management - Payment-based enhancements - Social or community services engagement - Team-based care - Technology enhancements	Population health

Study type	Author (Year)	Program name	Setting/context	Policy/ government program influencing innovation	Study design (quality evaluation rating)[a]	Patient population (if any)	Innovation elements included in the full intervention[b]	Types of outcomes studied
Observational cohort or cross-sectional study	Phillips et al (2014) [29]	The Illinois Medicaid Health Connect and Your Healthcare Plus programs	Illinois Medicaid beneficiaries, corresponding to 15% of the total state population	The Memisovski v. Maram suit (2004) ruled that Illinois had violated federal law by not providing adequate access to PC services for its Medicaid population, which made Illinois an early leader in Medicaid reform	Analysis of Medicaid claims and enrollment data from 2004 to 2010, covering both pre- and post-implementation (good, classified as retrospective cohort for quality evaluation)	Medicaid beneficiaries	- Accountability mechanisms - Care plan development - Care management - Improved access - Payment-based enhancements - Provider education or training - Technology enhancements	HC costs and utilization
Observational cohort or cross-sectional study	Wensing et al (2017) [30]	GP[j]-centered care (GPCC) program	Introduction of a program to enhance the role of general practice for patients with chronic diseases in Baden-Wuerttemberg, a German federal state with about 10.7 million inhabitants.	—	Comparative evaluation based on two cross-sectional studies at 4 and 5 years after its start (T1 and T2, respectively), based on data continuously collected for administrative control and reimbursement purposes (good)	General population aged 18 years or older with at least one primary care visit	- Accountability mechanisms - Efforts to improve performance monitoring - Enhanced coordination/ information exchange efforts - Improved access - Improved patient self-management - Inclusion of new/enhanced roles - Payment-based enhancements - Pharmacy/medication-related efforts - Provider education or training - Team-based care - Technology enhancements	HC costs and utilization
Case-control study	Freytag et al (2016) [31]	GP-centered program	A major Statutory Health Insurance fund AOK PLUS[k], which covers 41% of the population in central Germany, established a GP-centered health care program in 2011 in the German federal state of Thuringia	In Germany, enhanced primary care programs started in 2004 with the creation of a legal framework to support “GP-centered health care”	Retrospective case-control study based on insurance claims data (fair)	General patient population	- Inclusion of new/enhanced roles - Payment-based enhancements - Pharmacy/medication-related efforts - Provider education or training - Technology enhancements	HC costs and utilization

Study type	Author (Year)	Program name	Setting/context	Policy/ government program influencing innovation	Study design (quality evaluation rating)[a]	Patient population (if any)	Innovation elements included in the full intervention[b]	Types of outcomes studied
Pre-post study with no control	Conrad et al (2016) [32]	Group Health Cooperative's Access Initiative	PC practices within the integrated care delivery system that serves the Puget Sound region in Washington state	—	Pre-post implementation productivity assessment (good)	Group health cooperative's enrollees	- Enhanced service capacity - Improved access - Improved specialty care access - Others - Payment-based enhancements - Technology enhancements	HC costs and utilization
Pre-post study with no control	Engel et al (2016) [33]	Geriatrics in Primary Care (GPC)	Two large medical center practices at the Veterans Affairs Boston Healthcare System in 2014	Adoption of the Patient Aligned Care Team model of care, which is adapted from the PCMH, by the Veterans Affairs	Before-after evaluation of chart reviews (poor)	Veterans from the Veterans Affairs health system Boston, enrolled in the program	- Case/care management - Enhanced continuity/transition-based efforts - Enhanced service capacity - Improved access - Team-based care - Technology enhancements	HC costs and utilization
Pre-post study with no control	Maeng et al (2012) [34]	Proven-Health Navigator	36 Geisinger-owned PC practices, and seven contracted PC practices in GHP's provider network. Geisinger's regional health care system is a provider to regions of Pennsylvania and New Jersey	PCMH transformation in primary care.	Pre-post (measured at six points) and member fixed-effects model to measure within-member variation in the total cost and the PHN exposure variable over time (good)	GHP's Medicare Advantage plan members who were at least 65 years and enrolled in clinics that became PHN sites	- Case management - Efforts to improve performance monitoring - Enhanced service capacity - Improved access - Improved patient self-management - Payment-based enhancements - Social or community services engagement - Team-based care - Technology enhancements	HC costs and utilization

Study type	Author (Year)	Program name	Setting/context	Policy/ government program influencing innovation	Study design (quality evaluation rating)[a]	Patient population (if any)	Innovation elements included in the full intervention[b]	Types of outcomes studied
Pre-post study with no control	Ralston et al (2009) [35]	Group Health's Access Initiative	Adult respondents (aged ≥18 years) receiving care in Group Health's Western Washington Integrated Delivery System	Patient-centered system reforms (such as the PCMH model of 2007) mentioned as a shift in the way access to PC is provided, which encouraged HMOs[l] to change their restrictive access system.	Program impact evaluation, evaluating at three time points, based on the implementation dates of the initiative's components (fair)	Adult respondents (aged ≥18 years) receiving care in Group Health's Western Washington Integrated Delivery System	- Accountability mechanisms - Improved access - Improved specialty care access - Others - Payment-based enhancements - Technology enhancements	HC costs and utilization Patient satisfaction Provider satisfaction

[a] Ratings: good/fair/poor. Study type linked to the tool used for quality evaluation.

[b] Full details of innovation elements are provided in Multimedia Appendix 2.

[c] Not available or not reported in the articles.

[d] HC: health care.

[e] PC: primary care.

[f] T2DM: type 2 diabetes mellitus.

[g] GHP: Geisinger Health Plan.

[h] PCMH: Patient-Centered Medical Home.

[i] PHN: PatienHealthNavigator.

[j] GP: general practitioner.

[k] AOK PLUS: health insurance scheme under Germany insurer AOK.

[l] HMO: health maintenance organization.

In terms of study designs and quality evaluation results, three publications reported controlled interventions (two of "fair" and one of "good" quality), six reported observational cohort or cross-sectional studies with controls (four of "fair" and two of "good" quality), one reported a case-control study of "fair" quality, and four reported pre-post studies without controls (one of "poor", one of "fair," and two of "good" quality). Populations studied or linked to the results included the general population enrolled in the programs (in six articles), chronically ill patients with one disease or complex chronic patients (in four articles), and special populations, including elderly and disadvantaged populations (in four articles).

The interventions in the articles included between four and 11 "innovation features" (see Multimedia Appendix 2 for definitions). The average number of features per intervention was seven (median seven), and the most common types, beside technology-based enhancements (present in all interventions), were innovations to improve access (in 11 articles), payment-based enhancements (in nine articles), and care/case management (in seven articles). In terms of the types of outcomes, the most commonly reported was health care costs and utilization (in 10 articles), followed by population health outcomes (in four articles), patient satisfaction (in three articles), and provider satisfaction (in one article). These are not mutually exclusive as one article reported on three outcomes and two reported on two outcomes each. The remaining 11 articles reported on one outcome each.

Technology-Based Results

Of the 37 articles, 14 (38%) describing multicomponent interventions to enhance primary care included technology-based enhancements as one of the innovation elements (Table 2) [22-35].

Table 2. Technology types and details, aims, stakeholders involved, 4C support, and outcome summary (N=14).

Study	Technology-based on	Specific technology innovation	Aim and stakeholder (patient/provider/admin manager)	"4C" being supported by technology	General results and direction of the effects on quadruple health outcomes (of the full intervention)[a]
Coderch et al (2016) [22]	EMRs[b]	- Identification of patients: complex chronic patients are identified by labelling them in unique EMRs for providers - Proactive actions in PC[c]: individualized care plan registered in unique EMRs for providers	For providers, to be able to easily identify complex chronic patients under their care	Continuity	Health care costs and utilization ↑ (considerable increase in nonurgent primary care visits for partial and full interventions compared to each other and to control for both years 1 and 2) ↔ (mixed results for acute hospital admission and stay for year 1: considerable decrease for partial intervention compared to control, but considerable increase in full intervention compared to partial intervention; similar for readmissions <30 days in year 2, and considerable decrease for partial intervention and increase for full intervention when compared to each other) ↓ (increase in the number of prescriptions for full intervention compared to control for year 2)
Conrad et al (2008) [32]	Online messaging platform Online patient portal/website	- Patient-provider secure messaging through the MyGroupHealth enrollee website, including physician financial incentives for secure messaging patients - Internet access for enrollees to their EMRs through MyGroupHealth - Health promotion information on the MyGroupHealth secure website	For providers and patients, to have enhanced communication For patients, to promote self-management (through access to their medical information and health promotion information)	Continuity Comprehensiveness	Health care costs and utilization ↑ (considerable increase in panel size per FTE[d] and relative value unit per visit; considerable decrease in visits per FTE and per member per quarter costs) ↔ (nonrelevant increase in relative value unit per FTE)
Dale et al (2016) [25]	EMRs	- Optimal use of health IT[e], including improving EHR[f] function and capability and developing practice capability for optimal use of EHR; enabling exchange of patient information to support care; and developing quality measurement and reporting from EHRs	For providers, to better use EHRs, use information to support patient care, and improve quality monitoring	Comprehensiveness Coordination Continuity	Health care costs and utilization ↑ (decrease in total Medicare expenditures [without initiative care-management fees] and considerable decrease in PC visits and diabetes patients with no tests performed) ↔ (nonrelevant effects for hospitalizations, ED[g] visits, specialist visits, admissions for ambulatory care–sensitive conditions, and likelihood of readmissions; no differences for tests performed for diabetes or ischemic vascular patients) Patient satisfaction ↑ (increase in satisfaction with timely appointments, self-management support, and discussion of medications) ↔ (nonrelevant differences for communication with providers, knowledge of providers of other services, and patient ratings of providers)
Engel et al (2016) [33]	Telephone Electronic consultations	- Proactive telephone contact with veterans and caregivers, ready access to primary care colleagues, and informed use of telephone follow-up to enhance care while reducing nonessential clinic visits - Electronic consultation for formal referrals to geriatrics in PC program	For providers, to have easier referral to services For patients, to reduce clinical visits, while enhancing care	First contact Continuity Coordination	Health care costs and utilization ↑ (decrease in the number of specialist visits after years 1 and 2, while maintaining the number of PC visits)

Study	Technology-based on	Specific technology innovation	Aim and stakeholder (patient/provider/admin manager)	"4C" being supported by technology	General results and direction of the effects on quadruple health outcomes (of the full intervention)[a]
Freytag et al (2016) [31]	Medication-specific IT tool	- Obligatory use of a specific IT-pharmacotherapy tool to support rational pharmacotherapy	For providers, to support rational prescription of medicines	Comprehensiveness	Health care costs and utilization ↑ (decrease in the cost of drug prescriptions; increase in GP[h] consultations and decrease in specialist consultations, hospital use, and remedies; decrease in share of patients consulting more than one GP and accessing specialist without referrals; increase in the number of patients in disease management program and home visits; and decrease in the number of medical check-ups) ↓ (increase in the cost of GP consultations and specialist consultations and increase in the share of patients with five or more different medications) ↔ (no change in the number of ED hospitalizations or increase in the nursery care level)
Goff et al (2017) [26]	EHRs Use of insurer data	- Use of electronic health registries to identify patients in need of care and services (quarterly, reviewed the data contained in EHRs and insurer data focusing on specific care parameters in care [ie, ordered labs and mammography, scheduled PC visits, etc])	For providers, to monitor care needs and ensure tests and visits	Continuity	Population health ↑ (considerable changes in the mean DBP[i] and microalbumin/creatinine ratio test within 12 months) ↔ (no relevant difference for changes in HbA_{1c} measures, lipid measures, or other blood pressure measures; changes for HbA_{1c} tests and lipid panels)
Maeng et al (2012, 2012, 2013) [27,28,34]	EHRs Online patient portal Online messaging platform Modeling and utilization data tools	- Preventive and chronic care optimized by health IT. - Active delivery of information to other team members at the point of care via shared EHRs - Access to the patient portal for reviewing medical records and secure messaging with providers - Predictive modelling and utilization of data tools and normative management data to improve care	For providers, to have availability of patient information for all medical team members For providers and patients, to have enhanced communication For patients, to have access to their medical records to promote self-management For practices, to have improved monitoring for population care	Comprehensiveness Coordination Continuity	Population health ↑ (decrease in amputation and end-stage renal disease in the intervention group) ↔ (no difference for myocardial infarction or stroke) Health care costs and utilization ↑ (decrease in the per member and per month allowed costs; considerable overall savings with and without Rx coverage interaction) ↓ (increase in the cost of Rx coverage, without considering other program costs) Patient satisfaction ↑ (improvement in perceived changes in care delivery, ie, "noticed difference in care coordination and higher quality," increase in reporting of doctor's office as usual care, and decrease in ER[j] visits) ↔ (no relevant changes for access to care or primary care provider performance)
Phillips et al (2014) [29]	Online registries/report cards	- Multiple online tools, such as registries and report cards, to assist clinicians with population-based management	For providers, to have improved monitoring and population-based management	Continuity	Health care costs and utilization ↑ (increase in estimated cost savings and rate estimated annual savings; decrease in hospitalization, bed-day, and avoidable hospitalization rates; and increase in all quality measure changes [test and screenings]) ↔ (decrease in the ED visit rate for IHC[k] but increase for YHP[l])

Study	Technology-based on	Specific technology innovation	Aim and stakeholder (patient/provider/admin manager)	"4C" being supported by technology	General results and direction of the effects on quadruple health outcomes (of the full intervention)[a]
Prestes et al (2017) [23]	Data monitoring system	- The QUALIDIAB data system was used to verify the impact of the diabetes education intervention, and the data collected are useful to allocate resources (human and financial) considering real demand	For providers, to verify the impact of the intervention and allocate resources using collected data	Continuity	Population health ↑ (considerable improvements for DBP, glycemia, HbA_{1c}, total cholesterol, and LDL-c[m] and increase in the percentage of patients with target SBP[n] and HbA_{1c} levels) ↔ (nonrelevant differences for SBP, creatinine, proteinuria, HDL-c[o], DBP <80 mmHg, glycemia <100 mg/dL, cholesterol <200 mg/dL, and triglyceride <150 mg/dL) Health care costs and utilization ↑ (considerable increase in dyslipidemia patients treated, eye tests, and cardiovascular evaluations) ↔ (nonrelevant differences for dyslipidemia treated under target or any hypertension treatments)
Ralston et al (2009) [35]	Online patient portal Online messaging platform	- Web access for patients that provides secure email with physicians, medical record access, medication refills, appointment scheduling, discussion groups, and health promotion information	For patients, to facilitate accessing physicians, making appointments, refilling prescriptions, accessing medical records, and supporting self-management	First contact Comprehensiveness	Health care costs and utilization ↑ (improvement in "Getting Needed Care" and "Getting Care Quickly" scores) Patient satisfaction ↑ (improvement in satisfaction with the ability to see a personal doctor; time spent on the phone and waiting time for appointment; ease of getting care; and ratings of health care, health plan, and opinion of Group Health) Provider satisfaction ↑ (improvement in the perception of providers toward Group Health's quality and services provided and for Group Health as a good place to work)
Ruescas-Escolano et al (2014) [24]	EMRs	- Use of unique EMRs that allow for following control indicators and risk stratification	For providers, to monitor patient progress and manage risk	Continuity	Population health ↑ (considerable improvements in smoking status, cholesterol, and SBP) ↔ (nonrelevant differences for DBP)

Study	Technology-based on	Specific technology innovation	Aim and stakeholder (patient/provider/admin manager)	"4C" being supported by technology	General results and direction of the effects on quadruple health outcomes (of the full intervention)[a]
Wensing et al (2017) [30]	Medication-specific IT tool Updated IT systems	- The practice has a data-orientated quality system and decision support for prescribing medication; prompts in software to support use of generic and discounted drugs - The practice has up-to-date IT	For providers, to support medication prescription and promote generic medication use For practices, to have better organization to support easier patient access	Comprehensiveness First contact	Health care costs and utilization ↑ (decrease in the costs of medication therapy and hospital admissions) ↑ (increase in the number of visits to family physicians and mean number of prescription drugs; decrease in the number of prescriptions that should be avoided, contacts with specialists with and without referrals, hospital admissions, avoidable hospital admissions, number of days at hospital, and readmissions)

[a] Extracted from Jimenez et al [14].

[b] EMR: electronic medical record.

[c] PC: primary care.

[d] FTE: full-time equivalent.

[e] IT: information technology.

[f] EHR: electronic health record.

[g] ED: emergency department.

[h] GP: general practitioner.

[i] DBP: diastolic blood pressure.

[j] ER: emergency room.

[k] IHC: Illinois Medicaid Health Connect.

[l] YHP: Your Healthcare Plus.

[m] LDL-c: low-density lipoprotein cholesterol.

[n] SBP: systolic blood pressure.

[o] HDL-c: high-density lipoprotein cholesterol.

According to the descriptions of the articles, we were able to identify the following six broad categories for the types of implemented technologies (description below includes intended stakeholder and use):

1. 1. Enhancements leveraging *electronic medical/health records* [22,24-28,34]: it was the most common category (reported in seven studies from five interventions) and was aimed at providers. Their use is related to identifying specific groups of patients (eg, chronically ill) or specific needs of patients (eg, services needed), exchanging patient information, and developing quality measurements/control and risk stratification.
2. 2. *Data monitoring technologies/online registries* [23,26-29,34]: it was the second most common category (in six studies from four interventions) and was aimed at providers and practices. It was related to the management of utilization data to allocate resources and improve care, help with population-based management, and check on the impact of programs.
3. 3. Web-based *online portals and messaging platforms* [27,28,32,34,35]: it was included in five studies (reporting on three interventions) and was aimed at patients to access their medical records, obtain additional health promotion information, promote self-management, and facilitate access and communication with providers.
4. 4. *Medication-specific eHealth/information technology tools* [30,31]: it was included in two studies and was aimed at providers to support pharmacotherapy and medication prescription.
5. 5. *Telephone-based enhancements* [33]: it was described in one article and was aimed at providers to communicate with patients and caregivers, and provide follow-up to reduce patients' nonessential clinic visits.
6. 6. *Electronic consultations* between providers [33]: it was described in one study and was aimed at enhancing geriatric referrals.

Based on the description of the technological enhancements included in the studies, we were able to link them to the 4Cs in the following way:

1. 1. *First contact*: Three programs aimed to apply technology to impact this feature through telephone-facilitated access to primary care colleagues, facilitated appointment scheduling through web portals, and updated digital health systems for easier patient access.
2. 2. *Comprehensiveness*: Six interventions sought to increase the ability to manage a wider range of problems with technology, including providing additional health promotion information through patients' web portals and enhancing capacity for providers to better use electronic medical records, improve medication prescription, and provide improved preventive and chronic care.
3. 3. *Coordination*: Three programs used technology to improve care coordination by improving EHR-enabled information exchange and by allowing electronic

consultations to facilitate care among primary care providers and specialists.

4. 4. *Continuity*: Nine interventions sought to enhance the longitudinal relationship between patients and providers by enhancing the identification and follow-up of patients for individualized care, allowing more comprehensive identification and monitoring of service needs, and improving communication between patients and providers via online messaging or telephone contact.

Outcomes

Since these technology-based innovation elements are part of wider innovation environments, which include additional enhancement features, it was not possible to attribute outcomes specifically to the identified technologies. However, we still present the outcomes of the full innovation environments in an effort to elucidate the potential role of these technologies in the outcomes. The numerical magnitudes for each outcome are presented in Multimedia Appendix 3 (along with details of the full intervention). Table 2 and the paragraphs below present a descriptive summary and general direction of the effects for these outcomes.

Overall, the studies presented mixed results (ie, nonsignificant changes or significant benefits and deteriorations simultaneously for a specific outcome) for all types of outcomes, except for provider satisfaction, which was reported only in one study. The most consistent improvements per type of outcome were as follows: (1) *health care costs and utilization*, increased cost saving and decreased costs for some parameters (eg, Medicare expenditure decreased by US $11 per beneficiary per month [25] and drug prescriptions decreased by €44 per patient [31]) and increased primary care visits compared to specialists; (2) *population health*, improved blood pressure control, improved glycated hemoglobin, decreased amputations and end-stage renal disease, and decreased smoking status; (3) *patient satisfaction*, increased satisfaction with timely appointments and self-management support and increased satisfaction with the ability to see the usual doctor; (4) *provider satisfaction*, improved perception toward place of work's quality and services provided, and its consideration as a good place to work.

The most consistent mixed results by the type of outcome were as follows: (1) *health care costs and utilization*, nonsignificant changes or simultaneous improvements and deteriorations depending on the study for hospital admissions, readmissions, and emergency department visits; (2) *population health outcomes*, nonsignificant changes for cholesterol and lipid levels, myocardial infarction, and stroke; (3) *patient satisfaction*, no differences for communication with providers and for primary care provider performance.

The most consistent deteriorations were found for some health care costs and utilization outcomes, such as increased number of prescriptions, increased costs for general practitioner (eg, intervention €27 more expensive than control per patient) and specialist (intervention €22 more expensive than control per patient) consultations [31], and increased costs for prescription coverage.

Discussion

Principal Findings

Only 38% of our identified multicomponent interventions that aimed at enhancing primary care included technology-based enhancements, highlighting the fact that technology has not played a major role in comprehensive efforts aimed at enhancing primary care. This is not surprising, as it has been widely acknowledged that innovation in health care has always been difficult [36], especially if it has involved digital or technological efforts [37-39].

Most of the included articles reported on health care costs and utilization outcomes, signaling that technology-based efforts are either aimed at decreasing costs and utilization or at least not increasing costs without contributing to other aspects of system success. In fact, the only considerable unintended consequences were increased costs for general practitioner and specialist visits, and increased costs of prescription coverage (in some studies), suggesting that introducing technologies in health care can lead to increased costs, as it has been consistently reported in the literature [40-42].

The most common technology identified within these efforts was EHR, which is also not surprising given the widespread advocacy for this technology [43,44], and it was aimed mainly at providers or practices to facilitate information exchange among them and improve monitoring efforts. The only identified technology aimed at patients was the deployment of online patient portals, where they can see their records, message their providers, and access additional health information mostly for health promotion, which is in line with the idea that patients are ever more active participants in their own health care [44,45].

When analyzing the interventions in terms of their impact on the 4Cs, the technologies implemented were mostly aimed at improving continuity by increasing the identification and follow-up of patients (with labels in EHRs and telephone communication), enhancing monitoring efforts for identifying care and service needs (also mostly through EHRs and online registries), and ensuring more constant communication between providers and patients via online messaging. This reflects the growing importance of continuity of care, which in the past has had weak evidence linked to its benefits, but was recently highlighted as important, especially with regard to its link to decreased mortality risk [46,47]. Technologies have been promoted to improve comprehensiveness by providing additional health promotion information for patients, improving the ability of providers to prescribe medications, reinforcing the ability of primary care providers to cover a broader number of issues themselves, and avoiding overreferring [47].

In terms of outcomes, the literature provides limited but useful information. For example, increased patient satisfaction with timeliness of care, scheduling, and better self-management support could be in part explained by the use of online patient portals. Such portals allow patients to schedule appointments, see their own medical records, and access additional prevention information. Increased primary care visits, relative to specialist visits, appear to result from innovations that enhance monitoring of services needed and follow-up of patients (identified through EHRs and/or by telephone follow-up). The introduction of medication-specific digital/information technology tools could be associated with differing impacts. While studies reported a decrease in the costs of drug prescriptions and medication therapies, they also reported an increase in the mean number of drugs prescribed, and it was also associated with more costly consultations (around €25 extra per consultation [31]).

In order for digital technologies to play a more prominent role in primary care enhancement efforts, there is first a need for a responsible policy to support their development and introduction [48]. For example, some of the primary care enhancement environments have included an explicit policy encouraging the introduction of technology or information technology initiatives as part of their efforts [25,31,49]. To make this happen successfully, the technology must be seen as a tool that provides needed functions in a way that is effective, humane, and sustainable. Here, the context in which the technology will be implemented must be considered. It is essential to engage relevant stakeholders to deeply understand their environment and capabilities so that the introduced technology will be truly useful, improve (or at least not disrupt) existing workflow, and have tangible value [37,50]. In order to establish value, there is a need for technologies to be linked to evidence-based positive outcomes, taking into account their potential to improve health outcomes, costs, and patient and provider satisfaction.

There are some limitations for this study. The nature of the search and the specific requirements for including studies (ie, those describing multicomponent interventions aimed at enhancing primary care, which provided numerical magnitudes for reporting quadruple aim outcomes) may have made us overlook other important technological innovations aimed at improving primary care that had qualitative assessments only or did not measure quadruple aim outcomes. Similarly, since this review only focused on published and grey literature, it did not account for quality improvement interventions implemented at, for example, private primary care or accountable care organizations, which may have included technological innovations but no published results. Therefore, although technology may have a more prominent role in primary care–enhancing initiatives overall, the results from published and grey literature do not indicate so. Additionally, the fact that technological enhancements were one of many components within a primary care enhancement effort, our study eligibility criteria did not allow us to establish the actual and specific impact of the technologies on outcomes. However, it did help to situate these technologies within multicomponent innovation strategies and to gain preliminary insights into how technological enhancements may support other nontechnologically based innovation features and their impacts on the four primary care functions.

Taking all of this into account, future research should try to pinpoint the specific impact of technology-based innovation features within wider efforts aimed at enhancing primary care. This would mean including specific measures that could link outcomes to the use of such technology and quantify this impact. This exercise would also help to identify which of the 4Cs of primary care is being impacted by this technology, which would help understand the mechanisms of how these innovations are improving care. An additionally interesting research direction would be to explore how technological innovations are being leveraged at primary care private practices and/or accountable care organizations, to understand the actual role of technology in quality improvement initiatives for which there is no publicly available data. Such research could provide a more balanced view of the actual usage of technological innovations in primary care at the ground level.

Conclusions

Although technology and digital health have been proposed and encouraged as possible solutions to improve primary care, they have not played a major role in formally evaluated multicomponent interventions aimed at enhancing primary care, as reflected in the published and grey literature. Other types of nontechnologically based innovations, such as those aimed at improving access, restructuring payments for providers, and providing team-based care, have been much more widely implemented, reflecting that digital health technologies have not yet reached maturity or wider acceptance as a means for improving primary care. Leveraging technologies already in use, such as EHRs, and internet-based technologies, such as online patient portals, seems to provide promising avenues to improve continuity and comprehensiveness in primary care, which may eventually lead to better health outcomes and improved patient satisfaction. A stronger push is needed if technologies are meant to support wider efforts aimed at enhancing primary care and for them to play a more substantial role within these efforts. High-level policy and financial support must be designed to focus on the needs of a diversity of stakeholders and to encourage evidence-based research based on a coherent set of methods and measures. In this way, we can hope to fulfill the promise of technologies and digital health to enhance health care through strong sustainable primary care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[DOCX File , 14 KB - jmir_v23i1e20195_app1.docx]

Multimedia Appendix 2

Innovation elements and definitions.
[DOCX File , 16 KB - jmir_v23i1e20195_app2.docx]

Multimedia Appendix 3
Details of interventions and magnitudes of outcomes.
[DOCX File , 71 KB - jmir_v23i1e20195_app3.docx]

References

1. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005 Sep;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
2. Macinko J, Starfield B, Shi L. The contribution of primary care systems to health outcomes within Organization for Economic Cooperation and Development (OECD) countries, 1970-1998. Health Serv Res 2003 Jun;38(3):831-865 [FREE Full text] [doi: 10.1111/1475-6773.00149] [Medline: 12822915]
3. Haseltine W. Aging Populations Will Challenge Healthcare Systems All Over The World. Forbes Magazine. 2018. URL: https://www.forbes.com/sites/williamhaseltine/2018/04/02/aging-populations-will-challenge-healthcare-systems-all-over-the-world/?sh=1f071e852cc3 [accessed 2020-12-17]
4. Mendelson DN, Schwartz WB. The effects of aging and population growth on health care costs. Health Aff (Millwood) 1993 Jan;12(1):119-125. [doi: 10.1377/hlthaff.12.1.119] [Medline: 8509013]
5. Adler-Waxman A. This is the biggest challenge to our health. World Economic Forum: Agenda. 2017. URL: https://www.weforum.org/agenda/2017/12/healthcare-future-multiple-chronic-disease-ncd [accessed 2001-12-18]
6. Macinko J, Montenegro H, Nebot Adell C, Etienne C, Grupo de Trabajo de Atención Primaria de Salud de la Organización Panamericana de la Salud. [Renewing primary health care in the Americas]. Rev Panam Salud Publica 2007;21(2-3):73-84. [doi: 10.1590/s1020-49892007000200003] [Medline: 17565795]
7. A vision for primary health care in the 21st century: towards universal health coverage and the Sustainable Development Goals. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/328065 [accessed 2019-12-18]
8. Realising the Full Potential of Primary Health Care - Policy Brief. OECD Health Policy Studies. 2019. URL: http://www.oecd.org/health/health-systems/OECD-Policy-Brief-Primary-Health-Care-May-2019.pdf [accessed 2019-12-18]
9. Digital technologies: shaping the future of primary health care. World Health Organization. 2018. URL: https://www.who.int/docs/default-source/primary-health-care-conference/digital-technologies.pdf [accessed 2019-12-18]
10. Mitchell M, Kan L. Digital Technology and the Future of Health Systems. Health Syst Reform 2019 Mar 25;5(2):113-120. [doi: 10.1080/23288604.2019.1583040] [Medline: 30908111]
11. How Could Digital Technology Make an Impact on Primary Care? The Medical Futurist. 2018. URL: https://medicalfuturist.com/digital-technology-make-an-impact-on-primary-care/ [accessed 2019-12-18]
12. Starfield B. Primary care: concept, evaluation, and policy. Oxford, United Kingdom: Oxford University Press; 1992.
13. Higgings JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Hoboken, New Jersey: John Wiley & Sons Ltd; 2008.
14. Jimenez G, Matchar D, Koh GCH, Car J. Multicomponent interventions for enhancing primary care: a systematic review. Br J Gen Pract 2020 Nov 30. [doi: 10.3399/bjgp20X714199] [Medline: 33257458]
15. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014 Nov 10;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
16. OpenGrey. URL: http://opengrey.org/ [accessed 2020-12-22]
17. What is a health technology? World Health Organization. 2019. URL: https://www.who.int/health-technology-assessment/about/healthtechnology/en/ [accessed 2019-12-18]
18. Study Quality Assessment Tools. National Heart Lung and Blood Institute and National Institutes of Health. 2018. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2019-12-01]
19. Cantrell A, Croot E, Johnson M, Wong R, Chambers D, Baxter SK, et al. Access to primary and community health-care services for people 16 years and over with intellectual disabilities: a mapping and targeted systematic review. Health Serv Deliv Res 2020 Jan;8(5):1-142. [doi: 10.3310/hsdr08050] [Medline: 31999419]
20. Koppen IJN, Kuizenga-Wessel S, Saps M, Di Lorenzo C, Benninga MA, van Etten-Jamaludin FS, et al. Functional Defecation Disorders and Excessive Body Weight: A Systematic Review. Pediatrics 2016 Sep 16;138(3):e20161417-e20161417 [FREE Full text] [doi: 10.1542/peds.2016-1417] [Medline: 27531145]
21. Hashad N, Tonna A, Perumal D, Stewart D. Antimicrobial Stewardship Program Implementation in the Gulf Cooperation Council States: A Systematic Review. Journal of Infection and Public Health 2020 Feb;13(2):324. [doi: 10.1016/j.jiph.2020.01.052]
22. Coderch J, Pérez-Berruezo X, Sánchez-Pérez I, Sánchez E, Ibern P, Pérez M, et al. [Assessment of the effectiveness of a proactive and integrated healthcare programme for chronic complex patients]. Gac Sanit 2018;32(1):18-26 [FREE Full text] [doi: 10.1016/j.gaceta.2016.07.014] [Medline: 27789050]

23. Prestes M, Gayarre MA, Elgart JF, Gonzalez L, Rucci E, Paganini JM, DIAPREM (DIAbetes Primary Care, Registry, EducationManagement). Improving diabetes care at primary care level with a multistrategic approach: results of the DIAPREM programme. Acta Diabetol 2017 Sep 17;54(9):853-861. [doi: 10.1007/s00592-017-1016-8] [Medline: 28624898]
24. Ruescas-Escolano E, Orozco-Beltran D, Gaubert-Tortosa M, Navarro-Palazón A, Cordero-Fort A, Navarro-Pérez J, et al. El estudio PROPRESE: resultados de un nuevo modelo organizativo en atención primaria para pacientes con cardiopatía isquémica crónica basado en una intervención multifactorial. Atención Primaria 2014 Jun;46:10-15 [FREE Full text] [doi: 10.1016/S0212-6567(14)70060-5] [Medline: 25262306]
25. Dale SB, Ghosh A, Peikes DN, Day TJ, Yoon FB, Taylor EF, et al. Two-Year Costs and Quality in the Comprehensive Primary Care Initiative. N Engl J Med 2016 Jun 16;374(24):2345-2356. [doi: 10.1056/nejmsa1414953]
26. Goff SL, Murphy L, Knee AB, Guhn-Knight H, Guhn A, Lindenauer PK. Effects of an enhanced primary care program on diabetes outcomes. Am J Manag Care 2017 Mar 01;23(3):e75-e81 [FREE Full text] [Medline: 28385028]
27. Maeng DD, Davis DE, Tomcavage J, Graf TR, Procopio KM. Improving patient experience by transforming primary care: evidence from Geisinger's patient-centered medical homes. Popul Health Manag 2013 Jun;16(3):157-163. [doi: 10.1089/pop.2012.0048] [Medline: 23405878]
28. Maeng DD, Graf TR, Davis DE, Tomcavage J, Bloom FJ. Can a patient-centered medical home lead to better patient outcomes? The quality implications of Geisinger's ProvenHealth Navigator. Am J Med Qual 2012 Aug 18;27(3):210-216. [doi: 10.1177/1062860611417421] [Medline: 21852292]
29. Phillips RL, Han M, Petterson SM, Makaroff LA, Liaw WR. Cost, utilization, and quality of care: an evaluation of illinois' medicaid primary care case management program. Ann Fam Med 2014 Sep 08;12(5):408-417 [FREE Full text] [doi: 10.1370/afm.1690] [Medline: 25354404]
30. Wensing M, Szecsenyi J, Stock C, Kaufmann Kolle P, Laux G. Evaluation of a program to strengthen general practice care for patients with chronic disease in Germany. BMC Health Serv Res 2017 Jan 21;17(1):62 [FREE Full text] [doi: 10.1186/s12913-017-2000-2] [Medline: 28109281]
31. Freytag A, Biermann J, Ochs A, Lux G, Lehmann T, Ziegler J, et al. The Impact of GP-Centered Healthcare. Dtsch Arztebl Int 2016 Nov 25;113(47):791-798 [FREE Full text] [doi: 10.3238/arztebl.2016.0791] [Medline: 28043322]
32. Conrad D, Fishman P, Grembowski D, Ralston J, Reid R, Martin D, et al. Access intervention in an integrated, prepaid group practice: effects on primary care physician productivity. Health Serv Res 2008 Oct;43(5 Pt 2):1888-1905 [FREE Full text] [doi: 10.1111/j.1475-6773.2008.00880.x] [Medline: 18662171]
33. Engel PA, Spencer J, Paul T, Boardman JB. The Geriatrics in Primary Care Demonstration: Integrating Comprehensive Geriatric Care into the Medical Home: Preliminary Data. J Am Geriatr Soc 2016 Apr 21;64(4):875-879. [doi: 10.1111/jgs.14026] [Medline: 27100583]
34. Maeng D, Graham J, Graf TR, Liberman JN, Dermes NB, Tomcavage J, et al. Reducing long-term cost by transforming primary care: evidence from Geisinger's medical home model. Am J Manag Care 2012 Mar;18(3):149-155. [Medline: 22435908]
35. Ralston JD, Martin DP, Anderson ML, Fishman PA, Conrad DA, Larson EB, et al. Group health cooperative's transformation toward patient-centered access. Med Care Res Rev 2009 Dec;66(6):703-724. [doi: 10.1177/1077558709338486] [Medline: 19549993]
36. Herzlinger RE. Why innovation in health care is so hard. Harv Bus Rev 2006 May;84(5):58-66, 156. [Medline: 16649698]
37. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
38. Asthana S, Jones R, Sheaff R. Why does the NHS struggle to adopt eHealth innovations? A review of macro, meso and micro factors. BMC Health Serv Res 2019 Dec 21;19(1):984 [FREE Full text] [doi: 10.1186/s12913-019-4790-x] [Medline: 31864370]
39. Williams R. Why is it difficult to achieve e-health systems at scale? Information, Communication & Society 2016 Jan 11;19(4):540-550. [doi: 10.1080/1369118x.2015.1118521]
40. Callahan D. Chapter 17: Health Care Costs and Medical Technology. In: Crowley M, editor. From Birth to Death and Bench to Clinic: The Hastings Center Bioethics Briefing Book for Journalists, Policymakers, and Campaigns. Garrison, New York: The Hastings Center; 2008:79-82.
41. Kumar R. Technology and healthcare costs. Ann Pediatr Cardiol 2011 Jan;4(1):84-86 [FREE Full text] [doi: 10.4103/0974-2069.79634] [Medline: 21677816]
42. Zane RD, Wiler JL. Embracing Technology to Save Primary Care. NEJM Catalyst 2018;4(4). [doi: 10.1056/CAT.18.0101]
43. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
44. Staff. 10 Biggest Technological Advancements for Healthcare in the Last Decade. Becker's Health IT. 2015 Sep 17. URL: https://www.beckershospitalreview.com/healthcare-information-technology/10-biggest-technological-advancements-for-healthcare-in-the-last-decade.html [accessed 2019-12-18]

45. Dineen-Griffin S, Garcia-Cardenas V, Williams K, Benrimoj SI. Helping patients help themselves: A systematic review of self-management support strategies in primary health care practice. PLoS One 2019 Aug 1;14(8):e0220116 [FREE Full text] [doi: 10.1371/journal.pone.0220116] [Medline: 31369582]
46. Maarsingh OR, Henry Y, van de Ven PM, Deeg DJ. Continuity of care in primary care and association with survival in older people: a 17-year prospective cohort study. Br J Gen Pract 2016 Jun 20;66(649):e531-e539. [doi: 10.3399/bjgp16x686101]
47. Berenson RA, Burton R. How Solid Is The Primary Care Foundation Of The Medical Home? Health Affairs Blog 2016 Mar 25 [FREE Full text] [doi: 10.1377/hblog20160325.054144]
48. Pacifico Silva H, Lehoux P, Miller FA, Denis J. Introducing responsible innovation in health: a policy-oriented framework. Health Res Policy Syst 2018 Sep 10;16(1):90 [FREE Full text] [doi: 10.1186/s12961-018-0362-5] [Medline: 30200985]
49. Carter R, Quesnel-Vallée A, Plante C, Gamache P, Lévesque JF. Effect of family medicine groups on visits to the emergency department among diabetic patients in Quebec between 2000 and 2011: a population-based segmented regression analysis. BMC Fam Pract 2016 Feb 29;17(1):23 [FREE Full text] [doi: 10.1186/s12875-016-0422-2] [Medline: 26924443]
50. Swinkels ICS, Huygens MWJ, Schoenmakers TM, Oude Nijeweme-D'Hollosy W, van Velsen L, Vermeulen J, et al. Lessons Learned From a Living Lab on the Broad Adoption of eHealth in Primary Health Care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]

Abbreviations

EHR: electronic health record

Edited by G Eysenbach; submitted 13.05.20; peer-reviewed by M Reynolds, S Goff, N Taneja, P Rane, MDG Pimentel, M Rauws, F Palmieri, J Li; comments to author 08.07.20; revised version received 23.07.20; accepted 11.11.20; published 11.01.21.

Please cite as:
Jimenez G, Matchar D, Koh CHG, van der Kleij R, Chavannes NH, Car J
The Role of Health Technologies in Multicomponent Primary Care Interventions: Systematic Review
J Med Internet Res 2021;23(1):e20195
URL: http://www.jmir.org/2021/1/e20195/
doi:10.2196/20195
PMID:33427676

Review

Applications and Recruitment Performance of Web-Based Respondent-Driven Sampling: Scoping Review

Yannick B Helms[1,2], MSc; Nora Hamdiui[1,3], MSc; Mirjam E E Kretzschmar[2], Prof Dr; Luis E C Rocha[4], Prof Dr; Jim E van Steenbergen[1,5], MD, PhD; Linus Bengtsson[6], MD, PhD; Anna Thorson[7], MD, Prof Dr; Aura Timen[1,8], MD, Prof Dr; Mart L Stein[1], PhD

[1]National Coordination Centre for Communicable Disease Control, Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, Netherlands

[2]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

[3]Department of Primary and Community Care, Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, Netherlands

[4]Department of Economics & Department of Physics and Astronomy, Ghent University, Ghent, Belgium

[5]Centre for Infectious Diseases, Leiden University Medical Centre, Leiden, Netherlands

[6]Flowminder Foundation, Stockholm, Sweden

[7]Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden

[8]Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Yannick B Helms, MSc
National Coordination Centre for Communicable Disease Control
Centre for Infectious Disease Control
National Institute for Public Health and the Environment
Antonie van Leeuwenhoeklaan 9
Bilthoven
Netherlands
Phone: 31 30 274 70 00
Email: yannick.helms@rivm.nl

Abstract

Background: Web-based respondent-driven sampling is a novel sampling method for the recruitment of participants for generating population estimates, studying social network characteristics, and delivering health interventions. However, the application, barriers and facilitators, and recruitment performance of web-based respondent-driven sampling have not yet been systematically investigated.

Objective: Our objectives were to provide an overview of published research using web-based respondent-driven sampling and to investigate factors related to the recruitment performance of web-based respondent-driven sampling.

Methods: We conducted a scoping review on web-based respondent-driven sampling studies published between 2000 and 2019. We used the process evaluation of complex interventions framework to gain insights into how web-based respondent-driven sampling was implemented, what mechanisms of impact drove recruitment, what the role of context was in the study, and how these components together influenced the recruitment performance of web-based respondent-driven sampling.

Results: We included 18 studies from 8 countries (high- and low-middle income countries), in which web-based respondent-driven sampling was used for making population estimates (n=12), studying social network characteristics (n=3), and delivering health-related interventions (n=3). Studies used web-based respondent-driven sampling to recruit between 19 and 3448 participants from a variety of target populations. Studies differed greatly in the number of seeds recruited, the proportion of successfully recruiting participants, the number of recruitment waves, the type of incentives offered to participants, and the duration of data collection. Studies that recruited relatively more seeds, through online platforms, and with less rigorous selection procedures reported relatively low percentages of successfully recruiting seeds. Studies that did not offer at least one guaranteed material incentive reported relatively fewer waves and lower percentages of successfully recruiting participants. The time of data collection was shortest in studies with university students.

Conclusions: Web-based respondent-driven sampling can be successfully applied to recruit individuals for making population estimates, studying social network characteristics, and delivering health interventions. In general, seed and peer recruitment may be enhanced by rigorously selecting and motivating seeds, offering at least one guaranteed material incentive, and facilitating adequate recruitment options regarding the target population's online connectedness and communication behavior. Potential trade-offs should be taken into account when implementing web-based respondent-driven sampling, such as having less opportunities to implement rigorous seed selection procedures when recruiting many seeds, as well as issues around online rather than physical participation, such as the risk of cheaters participating repeatedly.

(J Med Internet Res 2021;23(1):e17564) doi:10.2196/17564

KEYWORDS

respondent-driven sampling; webRDS; online sampling; public health; interventions; research methodology; hard-to-reach populations; probabilistic sampling

Introduction

Respondent-driven sampling (RDS) is a sampling method that leverages social networks for recruiting individuals from populations that lack a sampling frame. The method has often been used to sample hard-to-reach populations, such as men who have sex with men, people who use intravenous drugs, and individuals with a migration background [1,2].

RDS starts with a convenience sample of members of a target population. The initially recruited participants (known as *seeds*) then recruit individuals from their social network (known as *peers*). These recruitees, in turn, invite their own peers and so on, resulting in a series of waves of recruitment [2,3]. Usually, RDS utilizes a coupon system to track who recruits whom, and requires that participants self-report the size of their social network within the target population [3]. These data can be used in a statistical model to account for the nonrandom data collection. As such, under certain assumptions, RDS qualifies as a probability sampling method that can generate unbiased population estimates [4].

RDS has several applications besides generating population estimates. For example, data on links between individuals (obtained through tracking the recruitment process) allow for studying interactions within and between participants' social networks. Among other things, this allows studying the spread of diseases in populations [5]. Furthermore, RDS can be used for recruitment of individuals for the delivery of health interventions [6,7].

Recruitment through RDS traditionally requires physical face-to-face interactions between individuals. However, over the past decade a novel online variant of RDS, so-called web-based RDS, was introduced. This potentially brings several benefits over offline RDS [8]. In particular, internet-based recruitment may (1) provide easy access and anonymity for participants; (2) overcome time- and location-related barriers to recruitment; and (3) provide an efficient, less laborious, and logistically demanding medium for recruitment from the researcher's perspective [8-12]. However, web-based RDS also introduces challenges, such as selection bias resulting from differential access to the internet and problems with the credibility of online research [10].

Nevertheless, the application of web-based RDS, its potential benefits, and its drawbacks for recruiting individuals have not yet been studied. Therefore, we aimed to provide an overview of web-based RDS applications and to investigate factors related to its recruitment performance, by means of a scoping review. We are aware that the main purpose of typical RDS is to generate population estimations. However, since we focus on recruitment through web-based RDS, in this study, we were equally interested in reported experiences with using web-based RDS for the recruitment of individuals for the characterization of social networks and the delivery of interventions. We also highlight potential areas for future research on web-based RDS and formulate general recommendations for researchers interested in its application.

Methods

Study Design

A scoping literature review [13] was conducted to gain insights into the application and performance of web-based RDS. We chose to conduct a scoping review because our aims were primarily exploratory, in the sense that we intended to provide an overview of the work done with web-based RDS so far and to identify factors related to the recruitment performance of web-based RDS. Table 1 provides an overview of the web-based RDS terminology used in this review (partially adapted from [14]).

Table 1. Meaning of important web-based RDS terminology.

Terms	Meaning
Participant	An individual participating in a study or intervention.
Peers	A participant's social contacts, such as friends or family members.
Coupon	An invitation (eg, in the form of a URL) that a participant can send to his/her peers, from the same target population, to invite peers in the study/intervention. Coupons use unique identifier codes to link recruiters with their recruitees.
Peer recruitment	The process of participants recruiting their peers.
Seed	A member of the target population who is recruited by a researcher to initiate peer recruitment.
Recruiter	A participant who recruits a peer by sending a coupon.
Recruitee	An individual who receives a coupon from a recruiter and agrees to enroll in the study/intervention.
Recruitment tree	A visualization of the peer recruitment process, in which all recruiters and their recruitees are linked in chains.
Wave	The distance (the number of chain-links) between seeds and their recruitees, in which seeds are in wave 0, their recruitees in wave 1, and so on.
Equilibrium	Equilibrium is reached when the sample composition of selected key indicators (eg, age and gender) remains stable over successive waves. Equilibrium indicates that the sample has become independent of the initially selected seeds.
Recruitment options	The options that participants have to forward their coupons to their peers.
Incentive	The stimuli provided to an individual for participation (primary incentive) or for each individual recruited (secondary incentive) to stimulate peer recruitment. An incentive can be material (tangible, eg, a gift card) or nonmaterial (intangible, eg, anonymous survey results).
Incentive structure	If only a primary or secondary incentive is offered, this is referred to as a single incentive structure; if both are offered, this is referred to as a double incentive structure.
Recruitment performance measures	Measures for recruitment performance (eg, number of individuals recruited) used in this research.

Search Strategy and Article Selection

We searched PubMed, Web of Science, and Scopus for articles. First, a preliminary search was conducted in PubMed to gauge the quality and quantity of web-based RDS related articles and to identify keywords to formulate the search syntaxes.

The term *web-based RDS* was introduced in 2008. In order to ensure that potentially relevant articles from before the term was introduced were included, we set our search range as 2000 to 2019. The following search terms were included in the final search syntaxes (see Multimedia Appendix 1 for the full syntaxes used):

1. Study type: implementation, development, testing, adoption, pilot
2. Online: online, web-based, internet, internet-based
3. Recruitment strategy: respondent-driven, peer-driven, participant-driven, snowball, chain-referral
4. Study purpose: intervention, sampling, recruitment, referral

Two researchers (YH and MS) independently screened the titles and abstracts of all unique records identified. The full text of selected records were then screened by one researcher (YH), to apply the below eligibility criteria. After this, the remaining articles were critically reviewed by YH, MS, and NH, before being included.

Eligibility Criteria

We included peer-reviewed articles that described the use of web-based RDS for the recruitment of participants for research purposes (ie, for making population estimates or for studying social networks) or for health intervention delivery. Articles that at least reported the numbers of seeds, subsequent recruitees, and discussed barriers and facilitators to the application of web-based RDS were included. We excluded studies that combined online and offline RDS without reporting on both approaches separately; if a study reported online and offline RDS separately, the online component was included. As we meant to provide an overview of the applications and performance of web-based RDS, our study was not limited to any particular target population or geographical area. We excluded studies that were not available in English or Dutch.

Data Extraction and Analysis

A data extraction table was developed to collect and organize data. The table's topics were iteratively identified and selected based on (1) STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) RDS guidelines [15], (2) topics discussed in a literature review with a similar purpose and context (offline RDS) to this study [1], and (3) discussion between YH, MS, and NH. Additional topics included study design, main findings, recommendations for further research, and limitations.

We used the process evaluation of complex interventions framework [16] to analyze the application and recruitment performance of web-based RDS. This framework explains the outcomes of an intervention as a function of implementation characteristics, mechanisms of impact, and contextual factors. We adapted this framework to fit web-based RDS specifications. In this study, we viewed web-based RDS as an intervention with the purpose of recruiting individuals. We defined outcomes as web-based RDS recruitment performance, implementation characteristics as the seed selection and recruitment process, mechanisms of impact as mechanisms to stimulate peer recruitment, and context as the setting in which web-based RDS was conducted.

Figure 1 (adapted from [16]) shows the analytical framework integrated with a schematic representation of the web-based RDS recruitment process. Table 2 shows the extraction table's topics, and operationalized measures thereof, grouped by components of the process evaluation framework.

Analyses focused on uncovering factors that influenced recruitment performance, based on comparing implementation characteristics, peer recruitment and recruitment performance measures between included studies. Data were presented in a narrative fashion.

Figure 1. Analytical framework for web-based RDS recruitment performance (adapted from [16]).

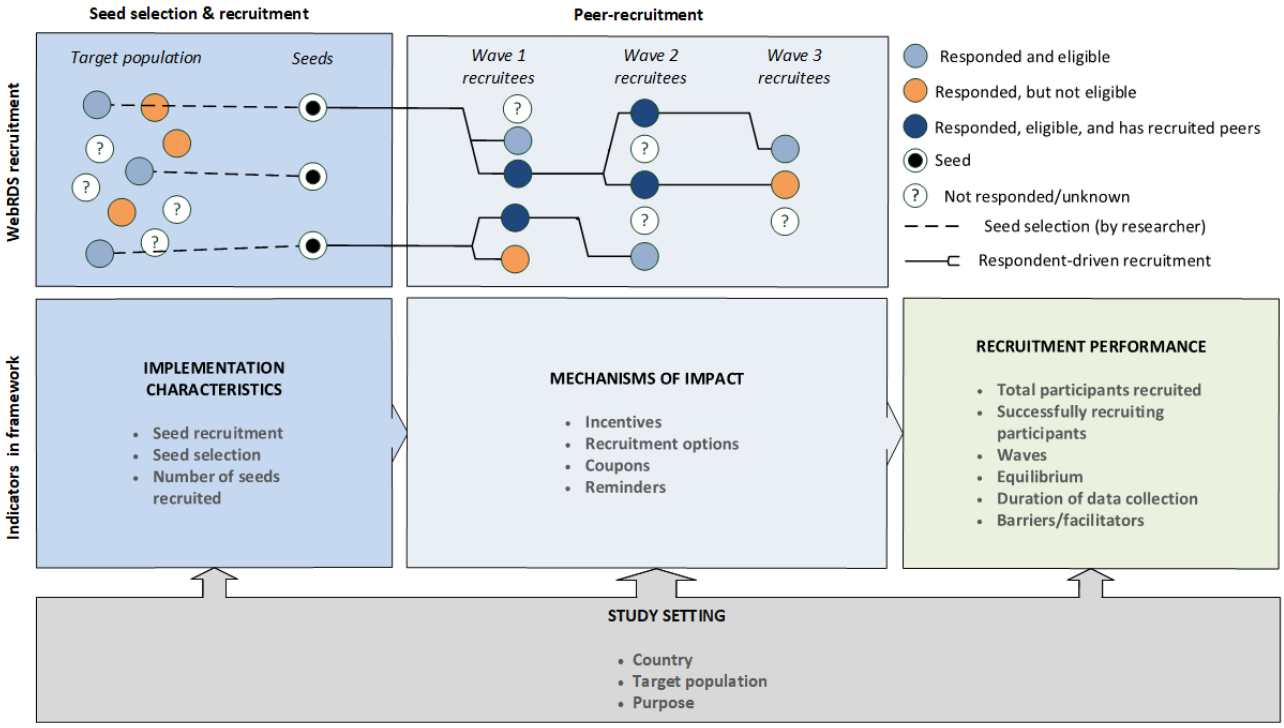


Table 2. Topics for data extraction and associated measures.

Topics	Measures
Study setting (context)	
Country	- Country
Target population	- Target population (specify)
Purpose	- Study purpose (eg, population estimates, social networks, intervention)
Seed recruitment and selection (implementation characteristics)	
Seed recruitment	- Recruitment platform (specify)
Seed selection	- Selection procedure (specify)
Number of seeds recruited	- Number of seeds recruited that participated in the study
Peer recruitment (mechanisms of impact)	
Incentives	- Material or nonmaterial (specify) - Single or double (specify) - Maximum value of incentive
Recruitment options	- Recruitment options (specify)
Coupons	- Number of coupons allowed (ie, that a participant can forward)
Reminders	- Reminders sent to enhance recruitment (yes/no)
Recruitment performance (outcomes)	
Total participants recruited	- Sample size[a]
Successfully recruiting participants[b,c]	- Proportion of seeds who successfully initiated recruitment - Proportion of all participants who successfully recruited peers to the study
Waves	- Maximum number of waves observed
Equilibrium	- Equilibrium reached (yes/no, and after how many waves)
Duration of data collection	- Duration of data collection
Barriers and facilitators	- Barriers and facilitators indicated to have influenced recruitment performance

[a] Sample size was calculated as sample size minus duplicate or fraudulent entries, if reported.

[b] We defined a successfully recruiting participant as a participant who invites at least one other person who participates in the study, regardless of the eligibility (the latter says more about how strict or elaborate researchers set eligibility criteria rather than about participants' ability to peer recruit). This excludes participants who merely sent out invitations with no response. If reported, duplicate or fraudulent entries were excluded.

[c] If not otherwise reported, this metric was manually counted and calculated from the reported recruitment tree.

Ethical Considerations

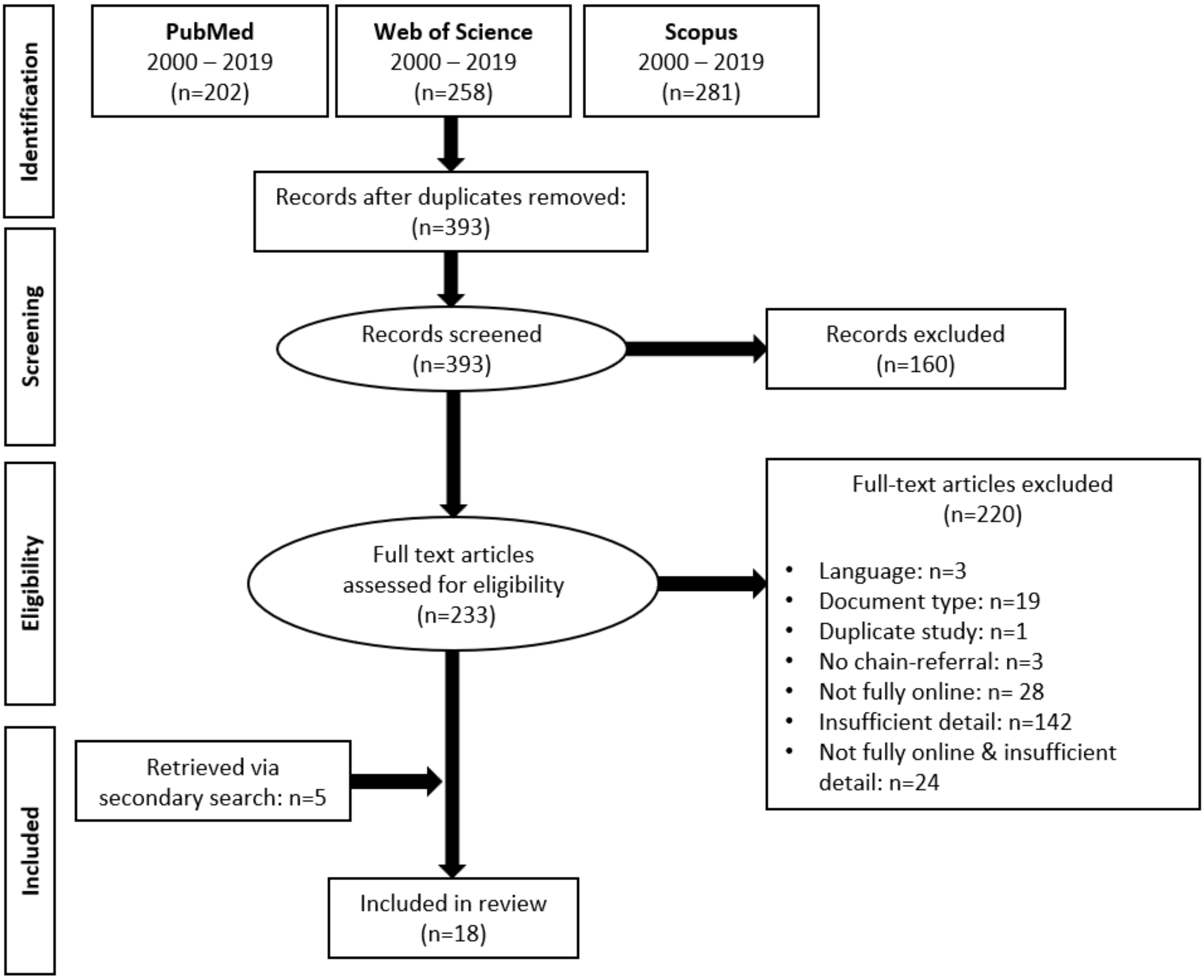
No ethical issues were foreseen in this study.

Results

Included Studies

We identified 393 unique records. The final number of articles included in this review was 18. See Figure 2 for a detailed account of the study inclusion procedure.

Figure 2. PRISMA flowchart.



Study Settings

Six studies took place in the United States [8,17-21], 6 studies took place in Western Europe [22-27], 4 studies took place in southeast Asia [23,28-31], and 2 studies took place in Oceania [32,33]. Studies were published between 2008 and 2019, with the majority [17,18,22-24,26,27,30,32,33] from 2015 onward (see Table 3).

In 12 studies, participants were recruited with the aim of generating population estimates [8,19-23,26,28,29,31-33]; 3 studies aimed to study social networks and contact patterns relevant to the spread of infectious diseases [24,25,28], and 3 studies recruited participants for delivering interventions [17,18,27].

Studies focused on a diverse set of target populations: the general population [19,24,25,28,30,32], university students [8,20], men who have sex with men [23,29,31], individuals who smoke [17], individuals using wheelchairs [33], parents of children aged 10 to 14 [18], people with precarious employment [22], young adults at risk of chlamydia infection [27], individuals who have migrated from Syria [26], and individuals who use marijuana [21].

Table 3. Characteristics of the articles included in review.

Reference	Setting	Study purpose	Target population	Final sample size (excluding fraudulent/duplicate entries, as reported)
Jonsson et al [22]	Stockholm county, Sweden	Generating population estimates	People with precarious employment	595 (551)
Weinmann et al [26]	Munich, Germany	Generating population estimates	Individuals who have migrated from Syria	195 (—[a])
Oesterle et al [18]	Washington and Colorado, United States	Intervention delivery	Parents of children in grades 6-8 (ie, aged 10-14 years)	235 (—)
Tran et al [30]	Vietnam	Generating population estimates	General population (youth and young adults)	366 (356)
Sadasivam et al [17]	United States	Intervention delivery	Individuals who smoke	759 (—)
Bourke et al [33]	New Zealand	Generating population estimates	Individuals using wheelchairs	19 (19)
Hildebrand et al [32]	Perth metropolitan area, Australia	Generating population estimates	General population (youth)	780 (—)
Stein et al [24]	Netherlands	Studying social networks and contact patterns relevant to the spread of infectious diseases	General population	1448 (1429)
Stromdahl et al [23]	Sweden	Generating population estimates	Men who have sex with men	148 (130)
Theunissen et al [27]	South Limburg, Netherlands	Intervention delivery	Young adults at risk of a *Chlamydia trachomatis* infection	68 (68)
Bengtsson et al [31]	Vietnam	Generating population estimates	Men who have sex with men	982 (870)
Crawford [21]	Oregon, United States	Generating population estimates	Individuals who use marijuana	72 (—)
Stein et al [25]	Netherlands and Thailand	Studying social networks & contact patterns relevant to the spread of infectious diseases	General population	358 (—)
Stein et al [28]	Thailand	Studying social networks & contact patterns relevant to the spread of infectious diseases	General population	257 (245)
Bauermeister et al [19]	United States	Generating population estimates	General population (young adults)	3448 (—)
Bengtsson et al [29]	Vietnam	Generating population estimates	Men who have sex with men	676 (591)
Wejnert [20]	Cornell University, United States	Generating population estimates	University students	378 (—)
Heckathorn and Wejnert [8]	Cornell University, United States	Generating population estimates	University students	159 (—)

[a] Not specified.

Recruitment-Related Results

Seed Recruitment and Selection

Five studies recruited seeds through online platforms in the form of targeted Facebook advertisements [17-19,32] or online participatory research panels [24] (see Multimedia Appendix 2), while 6 studies combined online (eg, online advertisements) with offline platforms, such as interest groups [23,26,28], researchers' social networks [22,25,26,28], or social venues [22,26,29,31] and 2 studies only recruited seeds offline—1 at a sexual health clinic [27] and 1 through a previous research project [21].

In most studies, researchers established contact with potential seeds as part of the seed selection procedure. This was done to confirm potential seeds' identity or eligibility [22,27,32,33], to select seeds with specific characteristics [19,22,26,30], or to confirm potential seeds' motivation and ability to invite and recruit peers [21-23,30]. Contact between researchers and seeds consisted of phone calls [19,22], emails [8,18,25,28], or in-person or group meetings [25,27,28]. In 3 studies, seed selection consisted only of an online eligibility screener and consent form [17,24,30].

The number of recruited seeds ranged between 1 [21] and 1015 [24].

Peer Recruitment

Most studies allowed participants to recruit their peers (eg, by sharing a URL) through preferred means of communication, such as WhatsApp or Facebook [17,19,21,22,29-32] or email [8,20,33]. Some studies additionally provided participants with the opportunity to provide their peers' contact details to the researchers, after which they contacted participants' peers via email (see Multimedia Appendix 2) [22,24,25,28,29,31].

In 4 studies, there was no limit for the number of coupons that participants could forward [17,18,27,30]. In other studies, the limit was 3 coupons [8,20,26,32,33], 4 coupons [22-25,28,29,31], 5 coupons [21], or 10 coupons [19].

Most studies had a double-incentive structure [8,17-20,22,23,26,29,31-33]. The majority of studies used material incentives as opposed to [8,19,20,22,23,26,30,32,33] or in combination with [17,18,24,29,31] nonmaterial incentives, and 4 studies only used nonmaterial incentives [21,25,27,28].

Material incentives were electronic gift cards [19,22,23,26,32], phone credits [29-31], or lottery tickets [24,29,31,33]. Nonmaterial incentives included showing participants aggregated study results [21,29,31] or visual insights into the study's recruitment process (eg, anonymized recruitment trees) [17,21,24,25,28], allowing participants to donate material incentives [29,31], and delivering political or social messages to prospective participants [18,21].

The maximum value of material incentives that participants could earn ranged between US $12.45 in Vietnam [29,31] and US $115 in the United States [17], but 2 studies had no maximum incentive value, since their number of incentivized coupons was unlimited [18,30].

Nine studies reported having sent reminders [17,22,24,27-29,31-33].

Recruitment Performance

The final sample sizes ranged between 19 [33] and 3448 [19]. Three studies recruited less than 100 participants [21,27,33], and 6 studies recruited more than 600 participants [17,19,24,29,31,32] (see Multimedia Appendix 2).

The proportion of seeds and the proportion of all participants who successfully recruited ranged between 7.5% [18] and 100% [21], and between 9.2% [27] and 55% [8], respectively. The maximum number of waves ranged between 1 [33] to 29 [31], and 8 studies reported fewer than 10 waves [17,21,23-25,27,28,33]. RDS sample distribution reached equilibrium in 5 studies, after 1 to 11 waves [8,20,22,26,29,31]. Data collection took between 72 hours [8] and 7 months [25]. In most studies, data collection took between 2 months and 3 months [18-21,26,28,30].

Factors Influencing Recruitment Performance

Overall, studies that recruited more seeds relatively often used online platforms (such as Facebook or other web communities) for seed recruitment, selected seeds less rigorously (eg, with less or no contact between potential seeds and researchers), and recruited seeds from larger geographical areas (eg, the entire country as opposed to a municipal area). Studies that recruited fewer seeds relatively often did so through a combination of both online and offline, or only offline platforms, with more elaborate seed selection procedures, and in smaller geographical areas.

Studies that recruited more seeds, through online platforms, and with less rigorous selection procedures, reported relatively lower percentages of successfully recruiting seeds. For example, the 3 studies that recruited the most seeds reported between 7.6% and 24.7% successfully recruiting seeds [17,18,24], compared to 67.4% to 100% in the 3 studies with the fewest recruited seeds [8,20,21].

Studies that did not offer at least one guaranteed material incentive (ie, not lottery-based compensation) reached no more than 6 waves and reported between 9.2% and 38.9% successfully recruiting participants [21,25,27,28].

Authors of included studies suggested a lack of monetary incentives [21,25,28,33], a lack of different recruitment options [19,22], and cheating (in order to retrieve multiple incentives) by participants [8,19,29,32] as potential factors related to web-based RDS protocols that influenced recruitment performance. The incompatibility of questionnaires or recruitment options with mobile platforms [18,19,23,30], and recruitment emails being identified as spam [8,19,28] were suggested as technical difficulties influencing recruitment performance. Seeds' motivation (or a lack thereof) to initiate recruitment [22,28,32,33], stigma regarding the study subject [27,33], online connectedness of the target population [8,17,22,26,33], and differential access to the Internet [8,18,19,28-32], were suggested as psychological and structural characteristics of the target population influencing recruitment performance.

Discussion

Overview

This is the first review investigating the application and recruitment performance of web-based RDS; a novel online sampling method. We identified 18 articles that described the use of web-based RDS. Out of all studies, 12 recruited participants for making population estimates, 3 recruited participants to study social network characteristics (contact patterns relevant to the spread of infectious diseases), and 3 recruited participants to deliver interventions. Studies were conducted in 8 countries, including both high- and low-middle income countries, over 4 continents. Between 19 and 3448 participants were recruited from various populations, including some without a sampling frame, such as men who have sex with men. The heterogenous nature of the included studies (with respect to their aims and setup) made it difficult to compare their recruitment processes and to draw generalizable inferences regarding recruitment performance.

Principal Findings

We found that studies that recruited relatively more seeds, through online platforms, and with less rigorous selection procedures reported lower percentages of successfully recruiting seeds. The exact reasons for this observation remain unclear. However, we suggest that recruiting more seeds relatively limits the time and resources available to thoroughly prepare (ie, motivate and inform) and select seeds. In turn, this may limit seeds' motivation to initiate peer recruitment, or lead to less suited (eg, less socially connected) seeds being selected. Both of these factors are known to be important for inducing and sustaining seed and peer recruitment [34,35].

Studies that did not offer at least one guaranteed material incentive reached relatively lower percentages of successfully recruiting participant, and fewer waves. We thus suggest that such incentives are particularly important to sustain recruitment, as sampling waves increase (ie, monetary incentives appear to carry further than nonmonetary incentives). This is in agreement with wider offline RDS literature [34] and indicates that benefits of online recruitment for participants (eg, easy access for participants) and nonmonetary incentives do not render material incentives redundant if the primary aim is to generate recruitment waves.

However, some studies that recruited participants for interventions reported relatively low percentages of successfully recruiting participants, despite offering substantial monetary incentives. This indicates that online peer recruitment for interventions benefits (or suffers) from factors other than peer recruitment for research purposes. Potentially, peer recruitment for interventions depends more on participants' affinity toward an intervention (eg, related to intervention framing, packaging), or its anticipated or experienced outcomes. Note, however, that these findings are based only on few studies and require further research to substantiate.

The majority of studies took between 2 and 3 months to collect data. However, online recruitment was relatively faster in some studies. For example, one study [8] achieved their targeted sample size (N=150) in 72 hours, and another [19] recruited 3448 participants in 6 weeks. Both these study populations were composed of university students who may be particularly digitally literate and have extensive well-connected online networks. This finding indicates the importance of these factors when considering applying this online method.

Strengths and Limitations

One strength of this literature review was the wide search strategy. It provides an extensive overview of peer reviewed literature relevant for investigating web-based RDS peer-recruitment processes. Another strength was the application of the process evaluation framework, which offered a practical structure for investigating different factors influencing web-based RDS recruitment performance.

One limitation is that we excluded all articles not reporting on the recruitment process in sufficient detail and all studies not exclusively using online peer recruitment or reporting on online and offline peer recruitment separately. Some valuable contextual and comparative information between online and offline recruitment might therefore have been missed. For some crucial recruitment performance measures (eg, the percentage of successfully recruiting participants), we had to rely on a manual count of recruitment trees, since the original data sets were unavailable.

Practical Implications and Opportunities for Future Research Using Web-Based RDS

Based on this review, it remains difficult to assess how successful web-based RDS is at achieving the aims for which it is employed (ie, generating population estimates, studying social networks, delivering interventions). For example, only 5 out of 12 studies aiming to generate population estimates reported that the sample composition reached equilibrium. Several studies likely achieved equilibrium (estimated from reported sample size and observed number of waves), but did not report this as such. Studies that used web-based RDS for studying social networks or delivering interventions were mostly feasibility or implementation studies, making it difficult to assess how successful the online method is at reaching the endpoints. At this point in time, we believe that there are not enough studies to draw meaningful conclusions regarding the overall success of web-based RDS for generating population estimates, studying social networks, or delivering interventions. Nevertheless, web-based RDS may be a particularly suited recruitment method when random sampling techniques are not feasible, the target population is geographically dispersed or hidden (which is a challenge for offline sampling), and the target population is extensively connected online [8]. Therefore, despite the heterogenous nature of the studies included in this review, which limits the generalizability of the studies' recruitment processes and performance, we outline several recommendations for future research into, or using web-based RDS.

First, consistent with offline RDS literature, the results suggest that recruiting a relatively small and thoroughly selected group of seeds (to whom a significant amount of resources can be dedicated for motivational and informing purposes) and providing at least one guaranteed material incentive is the most successful strategy for generating a substantial number of waves [1,34]. As such, this is the preferred setup for studies aiming to reach equilibrium for population estimates. If this is not the primary objective, for example when recruiting individuals for studying network characteristics or delivering interventions, recruiting a larger number of seeds through less rigorous means and providing lower or nonmaterial incentives may be preferred.

Second, despite the limited number of studies that recruited participants for interventions in this review, some implications in this regard stand out. For example, one study [27] found that through web-based RDS, individuals could be reached for sexually transmitted disease testing who were not reached before through traditional sexual health services. Another study [17] similarly noted that with each successive wave, the proportion of not-ready-to-quit smokers in the sample increased. These findings indicate that web-based RDS recruitment is particularly interesting for interventions if the aim is to reach more reluctant, or previously unreached individuals. The challenge here is to adequately incentivize peer recruitment (as discussed under principal findings).

Third, the results indicate that providing multiple recruitment options and facilitating the use of mobile platforms for participation and recruitment may enhance web-based RDS recruitment performance. However, it remains largely unclear how these factors influence peer recruitment across different settings, or even within certain target populations. For example, as indicated also by several studies included in this review, differential access to mobile communications, or the internet in general, may impose barriers to peer recruitment to readily excluded members of a given population. In addition, online communication behavior and the types of digital communication platforms used may differ between different networks, which could affect even relatively well-connected individuals. This is exemplified by one study [32] included in this review that compared web-based RDS to traditional offline RDS and found that individuals with lower socioeconomic status were less likely to be recruited through web-based RDS. Besides socioeconomic status, other factors known from literature that influence access to or use of the internet (and may therefore also influence online peer recruitment) include sociodemographic (eg, age, gender), socioeconomic (eg, household income, educational attainment), social (eg, degree of isolation, political context), and personal (eg, self-efficacy, mental health) factors [36]. To account for these potential sources of bias, we suggest thorough exploration of the target population’s online networks and communication behaviors, in a formative research stage.

Finally, we recommend that researchers using web-based RDS follow STROBE-RDS guidelines when reporting their studies [15]. A number of studies did not consistently report the numbers of total distributed and returned coupons, the numbers of recruitment waves, the numbers of individuals collecting their incentives, and the numbers of recruitees by seeds. Similar gaps in reporting on offline RDS data have been found in a previous review [37]. In addition, we encourage researchers to report relevant recruitment performance measures, such as the percentages of successful recruiters or the average numbers of recruitees per participant. This information is crucial for studying how to optimize peer recruitment in the future.

Conclusions

We have given a comprehensive overview of web-based RDS, what it is used for, how it is applied, and what factors influence its recruitment performance. Web-based RDS can be successfully applied to recruit individuals for making population estimates, studying social networks, and delivering health interventions. Peer recruitment may be enhanced by rigorously selecting and motivating seeds, offering at least one guaranteed material incentive, and facilitating adequate recruitment options regarding target populations’ online connectedness and communication behavior. Potential trade-offs should be taken into account when implementing web-based RDS. Examples are recruiting many seeds and less opportunities for rigorous seed selection procedures, as well as issues around online rather than physical participation, such as risks of cheaters through repeated participation. Furthermore, we have demonstrated the feasibility of—and described outcome measures for—research into web-based RDS recruitment using a process evaluation approach. The main points discussed in this literature review provide researchers with guidelines on key aspects and technicalities to consider when planning their web-based RDS research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search syntaxes.

[PDF File (Adobe PDF File), 61 KB - jmir_v23i1e17564_app1.pdf]

Multimedia Appendix 2

Recruitment results.

[PDF File (Adobe PDF File), 144 KB - jmir_v23i1e17564_app2.pdf]

References

1. Malekinejad M, Johnston LG, Kendall C, Kerr LRFS, Rifkin MR, Rutherford GW. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. AIDS Behav 2008 Jul;12(4 Suppl):S105-S130. [doi: 10.1007/s10461-008-9421-1] [Medline: 18561018]
2. Heckathorn DD, Cameron CJ. Network sampling: from snowball and multiplicity to respondent-driven sampling. Annu Rev Sociol 2017 Jul 31;43(1):101-119. [doi: 10.1146/annurev-soc-060116-053556]
3. Heckathorn DD. Snowball versus respondent-driven sampling. Sociol Methodol 2011 Aug 01;41(1):355-366 [FREE Full text] [doi: 10.1111/j.1467-9531.2011.01244.x] [Medline: 22228916]
4. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS 2005 May;19 Suppl 2:S67-S72. [doi: 10.1097/01.aids.0000172879.20628.e1] [Medline: 15930843]

5. Wejnert C. Social network analysis with respondent-driven sampling data: a study of racial integration on campus. Soc Networks 2010 May 01;32(2):112-124 [FREE Full text] [doi: 10.1016/j.socnet.2009.09.002] [Medline: 20383316]
6. Valente TW. Network interventions. Science 2012 Jul 06;337(6090):49-53. [doi: 10.1126/science.1217330] [Medline: 22767921]
7. Valente TW. Putting the network in network interventions. Proc Natl Acad Sci U S A 2017 Sep 05;114(36):9500-9501 [FREE Full text] [doi: 10.1073/pnas.1712473114] [Medline: 28851836]
8. Wejnert C, Heckathorn DD. Web-based network sampling. Sociological Methods & Research 2008 Aug 01;37(1):105-134. [doi: 10.1177/0049124108318333]
9. Baltar F, Brunet I. Social research 2.0: virtual snowball sampling method using Facebook. Internet Research 2012 Jan 27;22(1):57-74. [doi: 10.1108/10662241211199960]
10. Wright K. Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. J Comput Commun 2006;10(3):A. [doi: 10.1111/j.1083-6101.2005.tb00259.x]
11. Latkin CA, Knowlton AR. Social network assessments and interventions for health behavior change: a critical review. Behav Med 2015;41(3):90-97 [FREE Full text] [doi: 10.1080/08964289.2015.1034645] [Medline: 26332926]
12. Duncan DF, White JB, Nicholson T. Using internet-based surveys to reach hidden populations: case of nonabusive illicit drug users. Am J Health Behav 2003;27(3):208-218. [doi: 10.5993/ajhb.27.3.2] [Medline: 12751617]
13. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
14. Johnston L. Module 4: introduction to HIV/AIDS and sexually transmitted infection surveillance. Introduction to Respondent Driven Sampling-Manual. 2013. URL: https://www.researchgate.net/publication/286242895_Introduction_to_Respondent_Driven_Sampling-Manual [accessed 2019-03-03]
15. White RG, Hakim AJ, Salganik MJ, Spiller MW, Johnston LG, Kerr L, et al. Strengthening the Reporting of Observational Studies in Epidemiology for respondent-driven sampling studies: "STROBE-RDS" statement. J Clin Epidemiol 2015 Dec;68(12):1463-1471 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.04.002] [Medline: 26112433]
16. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
17. Sadasivam RS, Cutrona SL, Luger TM, Volz E, Kinney R, Rao SR, et al. Share2Quit: online social network peer marketing of tobacco cessation systems. Nicotine Tob Res 2017 Mar 01;19(3):314-323 [FREE Full text] [doi: 10.1093/ntr/ntw187] [Medline: 27613918]
18. Oesterle S, Epstein M, Haggerty KP, Moreno MA. Using Facebook to recruit parents to participate in a family program to prevent teen drug use. Prev Sci 2018 May 7;19(4):559-569 [FREE Full text] [doi: 10.1007/s11121-017-0844-7] [Medline: 29116552]
19. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838 [FREE Full text] [Medline: 22846248]
20. Wejnert C. An empirical test of respondent-driven sampling: point estimates, variance, degree measures, and out-of-equilibrium data. Sociol Methodol 2009 Aug 01;39(1):73-116 [FREE Full text] [doi: 10.1111/j.1467-9531.2009.01216.x] [Medline: 20161130]
21. Crawford SS. Revisiting the outsiders: innovative recruitment of a marijuana user network via web-based respondent driven sampling. Social Networking 2014;03(01):19-31. [doi: 10.4236/sn.2014.31003]
22. Jonsson J, Stein M, Johansson G, Bodin T, Strömdahl S. A performance assessment of web-based respondent driven sampling among workers with precarious employment in Sweden. PLoS One 2019 Jan 10;14(1):e0210183 [FREE Full text] [doi: 10.1371/journal.pone.0210183] [Medline: 30629661]
23. Strömdahl S, Lu X, Bengtsson L, Liljeros F, Thorson A. Implementation of web-based respondent driven sampling among men who have sex with men in Sweden. PLoS One 2015 Oct 1;10(10):e0138599 [FREE Full text] [doi: 10.1371/journal.pone.0138599] [Medline: 26426802]
24. Stein ML, van Steenbergen JE, Buskens V, van der Heijden PGM, Koppeschaar CE, Bengtsson L, et al. Enhancing syndromic surveillance with online respondent-driven detection. Am J Public Health 2015 Aug;105(8):e90-e97. [doi: 10.2105/AJPH.2015.302717] [Medline: 26066940]
25. Stein ML, van Steenbergen JE, Buskens V, van der Heijden PGM, Chanyasanha C, Tipayamongkholgul M, et al. Comparison of contact patterns relevant for transmission of respiratory pathogens in Thailand and The Netherlands using respondent-driven sampling. PLoS One 2014 Nov 25;9(11):e113711 [FREE Full text] [doi: 10.1371/journal.pone.0113711] [Medline: 25423343]
26. Weinmann T, AlZahmi A, Schneck A, Mancera Charry JF, Fröschl G, Radon K. Population-based assessment of health, healthcare utilisation, and specific needs of Syrian migrants in Germany: what is the best sampling method? BMC Med Res Methodol 2019 Jan 07;19(1):5 [FREE Full text] [doi: 10.1186/s12874-018-0652-1] [Medline: 30616507]
27. Theunissen K, Hoebe C, Kok G, Crutzen R, Kara-Zaïtri C, de VN, et al. A web-based respondent driven sampling pilot targeting young people at risk for Chlamydia trachomatis in social and sexual networks with testing: a use evaluation. Int

J Environ Res Public Health 2015 Aug 20;12(8):9889-9906 [FREE Full text] [doi: 10.3390/ijerph120809889] [Medline: 26308015]
28. Stein ML, van Steenbergen JE, Chanyasanha C, Tipayamongkholgul M, Buskens V, van der Heijden PGM, et al. Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: a pilot study in Thailand. PLoS One 2014 Jan 8;9(1):e85256 [FREE Full text] [doi: 10.1371/journal.pone.0085256] [Medline: 24416371]
29. Bengtsson L, Lu X, Nguyen QC, Camitz M, Hoang NL, Nguyen TA, et al. Implementation of web-based respondent-driven sampling among men who have sex with men in Vietnam. PLoS One 2012 Nov;7(11):e49417 [FREE Full text] [doi: 10.1371/journal.pone.0049417] [Medline: 23152902]
30. Tran BX, Zhang MW, Le HT, Nguyen HD, Nguyen LH, Nguyen QLT, et al. What drives young Vietnamese to use mobile health innovations? implications for health communication and behavioral interventions. JMIR Mhealth Uhealth 2018 Nov 30;6(11):e194 [FREE Full text] [doi: 10.2196/mhealth.6490] [Medline: 30251704]
31. Bengtsson L, Lu X, Liljeros F, Thanh HH, Thorson A. Strong propensity for HIV transmission among men who have sex with men in Vietnam: behavioural data and sexual network modelling. BMJ Open 2014 Jan 15;4(1):e003526 [FREE Full text] [doi: 10.1136/bmjopen-2013-003526] [Medline: 24435887]
32. Hildebrand J, Burns S, Zhao Y, Lobo R, Howat P, Allsop S, et al. Potential and challenges in collecting social and behavioral data on adolescent alcohol norms: comparing respondent-driven sampling and web-based respondent-driven sampling. J Med Internet Res 2015 Dec 24;17(12):e285 [FREE Full text] [doi: 10.2196/jmir.4762] [Medline: 26704736]
33. Bourke JA, Schluter PJ, Hay-Smith EJC, Snell DL. Respondent driven sampling of wheelchair users: a lack of traction? F1000Res 2016 Apr 26;5:753. [doi: 10.12688/f1000research.8605.1]
34. Johnston LG, Sabin K. Sampling hard-to-reach populations with respondent driven sampling. Methodological Innovations Online 2010 Aug 01;5(2):38.1-3848. [doi: 10.4256/mio.2010.0017]
35. Johnston LG, Malekinejad M, Kendall C, Iuppa IM, Rutherford GW. Implementation challenges to using respondent-driven sampling methodology for HIV biological and behavioral surveillance: field experiences in international settings. AIDS Behav 2008 Jul;12(4 Suppl):S131-S141. [doi: 10.1007/s10461-008-9413-1] [Medline: 18535901]
36. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. a systematic review of the second- and third-level digital divide. Telematics and Informatics 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
37. Johnston LG, Hakim AJ, Dittrich S, Burnett J, Kim E, White RG. A systematic review of published respondent-driven sampling surveys collecting behavioral and biologic data. AIDS Behav 2016 Aug 18;20(8):1754-1776 [FREE Full text] [doi: 10.1007/s10461-016-1346-5] [Medline: 26992395]

Abbreviations

RDS: respondent-driven sampling
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

---

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by J Duggan, T Lam; comments to author 12.06.20; revised version received 26.06.20; accepted 19.07.20; published 15.01.21.

Please cite as:
Helms YB, Hamdiui N, Kretzschmar MEE, Rocha LEC, van Steenbergen JE, Bengtsson L, Thorson A, Timen A, Stein ML
Applications and Recruitment Performance of Web-Based Respondent-Driven Sampling: Scoping Review
J Med Internet Res 2021;23(1):e17564
URL: http://www.jmir.org/2021/1/e17564/
doi:10.2196/17564
PMID:33448935

Review

Implementation of Telehealth Services to Assess, Monitor, and Treat Neurodevelopmental Disorders: Systematic Review

Althea Z Valentine[1], BSc, MSc, PhD; Sophie S Hall[2], PhD; Emma Young[3], MSc; Beverley J Brown[1], PhD; Madeleine J Groom[1], PhD; Chris Hollis[1,4,5], PhD; Charlotte L Hall[1], PhD

[1]Institute of Mental Health, School of Medicine, Division of Psychiatry and Applied Psychology, University of Nottingham, Nottingham, United Kingdom

[2]Department of Neuroscience, Psychology and Behaviour, University of Leicester, Centre for Medicine, Leicester, United Kingdom

[3]Nottinghamshire Healthcare National Health Service Foundation Trust, Library and Knowledge Services, Duncan Macmillan House Staff Library, Nottingham, United Kingdom

[4]National Institute for Health Research MindTech MedTech Co-operative, Institute of Mental Health, School of Medicine, Division of Psychiatry and Applied Psychology, University of Nottingham, Nottingham, United Kingdom

[5]Department of Child and Adolescent Psychiatry, Queen's Medical Centre, Nottingham, United Kingdom

Corresponding Author:
Althea Z Valentine, BSc, MSc, PhD
Institute of Mental Health
School of Medicine, Division of Psychiatry and Applied Psychology
University of Nottingham
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 01157484249
Email: althea.valentine@nottingham.ac.uk

Abstract

Background: In response to COVID-19, there has been increasing momentum in telehealth development and delivery. To assess the anticipated exponential growth in telehealth, it is important to accurately capture how telehealth has been used in specific mental health fields prior to the pandemic.

Objective: This systematic review aimed to highlight how telehealth has been used with clinical samples in the neurodevelopmental field, including patients with neurodevelopmental disorders (NDDs), their families, and health care professionals. To identify which technologies show the greatest potential for implementation into health services, we evaluated technologies for effectiveness, economic impact, and readiness for clinical adoption.

Methods: A systematic search of literature was undertaken in April 2018 and updated until December 2019, by using the Medline, Web of Science, Scopus, CINAHL Plus, EMBASE, and PsycInfo databases. Extracted data included the type of technology, how the technology was used (ie, assessment, treatment, and monitoring), participant characteristics, reported outcomes and authors' views on clinical effectiveness, user impact (ie, feasibility and acceptability), economic impact, and readiness for clinic adoption. A quality review of the research was performed in accordance with the Oxford Centre for Evidence-Based Medicine Levels of Evidence.

Results: A total of 42 studies met the inclusion criteria. These studies included participants and family members with autism spectrum disorders (21/42, 50%), attention deficit hyperactivity disorders (8/42, 19%), attention deficit hyperactivity or autism spectrum disorders (3/42, 7%), communication disorders (7/42, 17%), and tic disorders (2/42, 5%). The focus of most studies (33/42, 79%) was on treatment, rather than assessment (4/42, 10%) or monitoring (5/42, 12%). Telehealth services demonstrated promise for being clinically effective, predominantly in relation to diagnosing and monitoring NDDs. In terms of NDD treatment, telehealth services were usually equivalent to control groups. There was some evidence of positive user and economic impacts, including increased service delivery efficiency (eg, increased treatment availability and decreased waiting times). However, these factors were not widely recorded across the studies. Telehealth was demonstrated to be cost-effective in the few studies that considered cost-effectiveness. Study quality varied, as many studies had small sample sizes and inadequate control groups. Of the 42 studies, only 11 (26%) were randomized controlled trials, 12 (29%) were case studies or case series, 6 (14%) were qualitative studies, and 5 (12%) were noncomparative trials.

Conclusions: Telehealth has the potential to increase treatment availability, decrease diagnosis waiting times, and aid in NDD monitoring. Further research with more robust and adequately powered study designs that consider cost-effectiveness and increased efficiency is needed. This systematic review highlights the extent of telehealth technology use prior to the COVID-19 pandemic and the movement for investing in remote access to treatments.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42018091156; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018091156

(J Med Internet Res 2021;23(1):e22619) doi:10.2196/22619

KEYWORDS

neurodevelopmental disorders; technology; telehealth; review; COVID-19; implement; effective; mental health

Introduction

Background

Neurodevelopmental disorders (NDDs) are lifelong disorders that typically develop during the early stages of child development and have a high frequency of co-occurrence [1,2]. In this systematic review, NDDs are defined in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria [3], and include autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), intellectual disability, communication disorders, specific learning disorder, motor disorders, stereotypical movement disorder, and tic disorders. Young people with NDDs have been identified as particularly vulnerable to the mental health impacts of COVID-19, due to changes in support and routine and increased isolation and loneliness [4,5].

Prior to the COVID-19 pandemic, telehealth interventions were attracting interest as effective options for improving mental health provision in overstretched health services. The COVID-19 pandemic has increased the demand for effective mental health support, and the growing need to offer easy-to-access remote service availability [4,6,7] has substantially increased telehealth use [8]. It is therefore essential that we not only identify which existing telehealth technologies show the greatest efficacy for use with individuals with NDDs, but also capture the state of the existing evidence base in order to evaluate the inevitable growth of this field.

Prior Work

There is no universally agreed upon definition for telehealth [9]. In this systematic review, we use the term “telehealth” to encompass telemedicine, telemental health, and telepsychiatry.

In a systematic review of the use of telehealth services for communication disorders, Molini-Avejonas and colleagues [10] found that over 85% (88/103) of telehealth studies reported the advantages that telehealth has over nontelehealth approaches. For example, Molini-Avejonas and colleagues [10] reported that telehealth is typically viewed favorably by users and health care practitioners, as telehealth helps to reduce geographical barriers and possibly save time during consultations and travel. However, barriers to telehealth implementation have been identified. These barriers relate to training, technology issues, and acceptance by both health care practitioners and patients [10]. Indeed, a study that explored the views of health care practitioners (ie, neurologists) toward digital devices in clinical practice found that while the majority (95%) of the 405 participants used computers regularly at work, less than half (43.5%) used a tablet [11]. This suggests that one of the barriers to the uptake of technology may be acceptance from health care professionals.

Sutherland and colleagues [12] have also updated a systematic review [13] of telehealth literature on participants with ASD. During 2010-2016, 14 studies with a total of 284 ASD participants assessed telehealth services, including assessments, interventions, functional behavioral analyses, and language therapy. These studies included a variety of controls, including comparisons between telehealth and face-to-face sessions (6/14, 43%), online learning with and without telehealth sessions (6/14, 43%), and telehealth services that provided no intervention and those that provided treatment as usual (2/14, 14%). Although these studies varied in quality, telehealth services were comparable to face-to-face services and better than control/comparison groups in experimental studies. Another systematic review found that telehealth systems have been used to deliver education to parents and support the diagnosis and treatment of ASD [14].

In terms of ADHD, only 1 systematic review has focused on the use of telehealth. This review found 11 articles, which all reported data from 3 trials that were conducted in 2007-2017 [15]. The majority (10/11, 91%) of studies used a sample of children. Telehealth was viewed favorably, as it was well accepted by health care professionals and users and shown to provide improved outcomes, such as reduced symptomology and improved functioning. However, the authors concluded that further research was necessary to assess the usefulness of telehealth in health care delivery [15]. This review highlighted a lack of research on using telehealth to replace usual treatment rather than augment usual treatment, and a lack of studies that consider the assessment, diagnosis, and treatment of adults with ADHD.

Although condition-specific systematic reviews have been conducted, no single review has assessed the use of telehealth across people with different NDDs. This is important, given the prevalence of NDD comorbidities. Many previous reviews have also been limited to trials. Although trials are important, user feedback, economic impact, and readiness for clinical adoption are important for rapidly developing policies for implementing telehealth services after the COVID-19 pandemic.

The Goal of This Study

The aim of this systematic review was to highlight how telehealth has been used, prior to the COVID-19 pandemic, with clinical samples within the neurodevelopmental field, including patients with NDD, their families, and health care professionals. In light of the post-COVID-19 pandemic call for implementing the rapid adoption of telehealth into clinical practice [16], this systematic review focused on studies that reported on the clinical/service effectiveness, economic impact, and user impact (ie, feasibility/acceptability) of telehealth to aid in assessment, diagnosis, monitoring, and treatment. This review serves to identify potentially effective telehealth technologies for use with patients with NDDs and document the evidence base prior to the anticipated rapid expansion of telehealth in the neurodevelopmental field.

Methods

Study Design

This systematic review was part of a larger review [17], which assessed all technology that has been used for NDDs. The protocol for our main review was registered with PROSPERO (CRD42018091156). Given the vast number of obtained papers that related to telehealth, it was most appropriate to present these in a stand-alone article. The literature search was undertaken in accordance with the recommended principles in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [18].

Search Strategy

A systematic search of literature was undertaken by an information specialist (EY) using the following databases: Medline, Web of Science, Scopus, CINAHL Plus, EMBASE, and PsycInfo. Searches were also performed in the Cochrane Library, Journal of Medical Internet Research, Institute of Electrical and Electronics Engineers, and Association for Computing Machinery Digital Library databases. The search included all terms that related to NDDs and telehealth, including controlled vocabulary headings such as “Intellectual Disability,” “Mentally Disabled Persons,” “Learning disorders,” “Developmental Disabilities,” “Neurodevelopmental Disorders,” and “Telemedicine.” Keywords and synonyms that related to all NDDs, including “ASD,” “ADHD,” “Tic Disorders,” “Communication and Language Disorders,” “Learning Disorders,” and “Learning Disabilities,” were also used for the search. Terms that related to telehealth included keywords, such as “tele care,” “tele coaching,” “telecomm,” “teleconference,” “teleconsultation,” “telehealth,” and “telemanagement,” as well as terms that related to teletherapy, telepractice, and eHealth. As this study was part of a wider search of all technologies, additional terms that related to various technologies, such as mobile apps, video games, virtual reality, and robotics, were also included. However, the results of the search for these terms are presented in another study [17]. A copy of the Medline search strategy is included in Multimedia Appendix 1. Endnote software (Clarivate) and Microsoft Excel were used to manage the data. The initial search was restricted to published, peer-reviewed, academic papers written in English, and was conducted in March/April 2018 and recently updated in July 2020 to cover the period of January 2014 to December 2019. The World Health Organization has acknowledged December 2019 as the month that the first case of COVID-19 was officially recorded [19].

PICOS (population, intervention, comparison, outcome, study design) guidelines were used to define the inclusion criteria. With regard to population, we included studies that involved people with NDDs or parents, carers, or health care professionals who worked with people with NDDs. With regard to intervention, we included studies that clinically used telehealth equipment in the assessment, diagnosis, monitoring, or treatment of NDDs. No restrictions on comparisons were put in place for literature. With regard to outcomes, included studies were to have at least 1 outcome of interest from clinical effectiveness, economic impact, and user impact. Based on the National Institute of Clinical Excellence glossary, the following terms were referred to in the search: (1) “clinical effectiveness,” which refers to how beneficial telehealth was in terms of assessment, monitoring, or treatment compared to usual care, a control group, or another type of care; (2) “economic impact,” which refers to the evaluation of service delivery efficiencies (eg, whether an intervention reduces clinician time), as well as any economic evaluation (eg, cost-effectiveness or costs and benefits evaluations) of telehealth; and (3) “user impact,” which refers to the feasibility of using telehealth in terms of technical feasibility (ie, how simple or difficult it was to use telehealth services) and the administrative infrastructure (ie, how the technology fits within an organization). Usability impact also covered design factors that affect the user experience and users’ acceptability of the technology (ie, users’ willingness to attend and engage with the technology). With regard to study types, we excluded systematic reviews and meta-analyses.

Studies on telehealth were restricted to those that used synchronous (ie, real-time) remote clinical care in relation to the diagnosis, monitoring, or treatment of an NDD. Although studies that involved both audio and video communication were included, studies that provided care via only a telephone were excluded. Studies were also excluded if they used asynchronous (ie, nonreal-time) data, including email communications between patients and health care practitioners, physiological data (eg, electroencephalogram data) that were remotely interpreted, and data regarding telehealth services that were delivered solely in educational/employment settings, such as schools or vocational training centers. In addition, studies were excluded if they did not involve an NDD clinical sample or if they focused on lifestyle interventions (eg, obesity management rather than NDD treatment).

Data Extraction and Quality Assessment

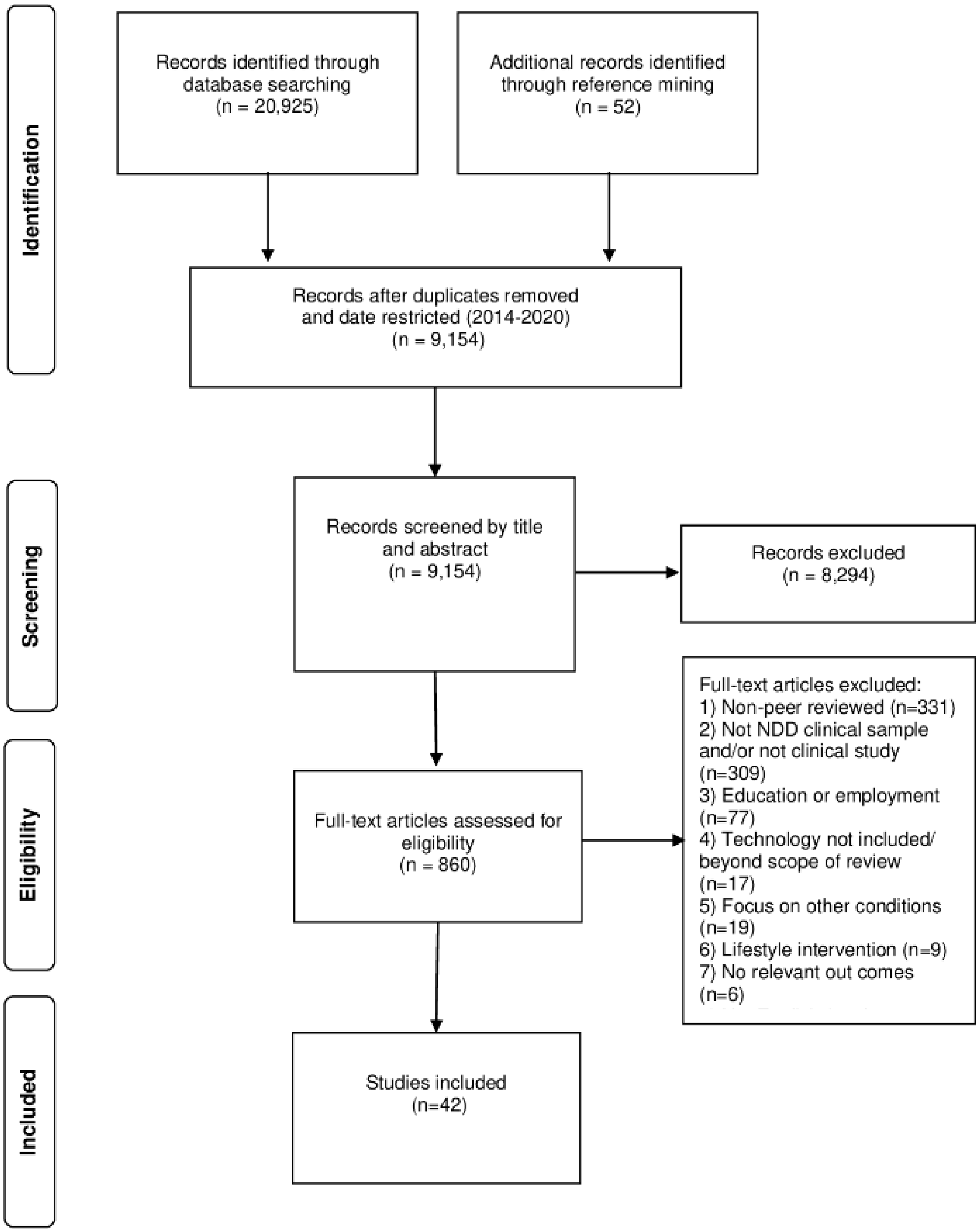
Titles and abstracts were reviewed for initial screening, and excluded papers were further independently screened. Two authors (AZV and CLH) independently reviewed full texts and extracted data by using an Excel database. Extracted data included authors and the year of publication; brief summaries of the study design, including the type of telehealth used and study methods; how the technology was used (ie, assessment, treatment, or monitoring); and information on participant samples, including the number of participants in a sample, health condition, gender, population type (ie, parent, clinician, or children/young people populations), and age (ie, if children were studied). The relevant outcomes that related to the authors' views on clinical effectiveness, user impact, economic impact, and readiness for clinic adoption were also noted. Results were synthesized in tabulated form (Multimedia Appendix 2).

A quality review of the research was also conducted. Papers were appraised by 3 authors (CLH, SSH, and BJB) based on the Oxford Centre for Evidence-Based Medicine Levels of Evidence. Each paper was rated with a score of 1-5; randomized controlled trials (RCTs) were typically ranked high (score=1) and qualitative papers/judgments were typically ranked low (score=5). Throughout the paper, this score is referred to as a quality rating (QR) [20]. Disagreements were resolved through discussion.

Results

The process of identifying and selecting studies is outlined in a flow diagram (Figure 1), and a summary of the included papers is presented in Multimedia Appendix 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 flow diagram depicting the study selection process.



Study Characteristics

A total of 42 studies met the inclusion criteria. The greatest number of studies were conducted on ASD (22/42, 52%) and ADHD (8/42, 19%). Studies on communication disorders (7/42, 17%), and tic disorders (2/42, 5%) were the least represented. Additionally, 3 (7%) studies used a sample of participants with ADHD, ASD, or both. Of the 42 papers, 23 (55%) reported a wide range of additional diagnoses, such as another coexisting NDD (10/23, 43%), oppositional defiant disorder (7/23, 30%), and anxiety (4/23, 17%), and 19 (45%) studies did not report any comorbidities. Most studies (29/42, 69%) reported data from children's parents/carers. Of these 29 studies, 22 (76%) included children aged <7 years, 3 (10%) included adult telehealth service users, and 7 (30%) documented the perspective of health care professionals. Approximately half (20/42, 48%) of all papers reported on data from a male or predominantly male sample. However, in parent/carer studies, the primary caregiver was more likely to be female. Most studies were conducted in the United States (27/42, 64%), Australasia (7/42, 17%) and Europe (6/42, 14%). Studies typically focused on treatment (33/42, 79%), rather than monitoring (5/42, 12%) or assessment (4/42, 10%).

Outcomes of interest

Assessment

Summary of Assessment Papers

We found 4 papers that focused on the assessment of NDDs. Of these 4 papers, 3 (75%) used telehealth to remotely diagnose ASD [21-23] and 1 (25%) assessed the objective measurement of hyperactivity in patients diagnosed with ADHD [24]. All ASD studies involved parents and children under 6 years of age. The ADHD study involved children and young people aged 6-16 years. All studies had ≤65 participants (range 17-65; Multimedia Appendix 2). Of the 4 papers, 3 (75%) had a QR of 2 [22] or 3 [23,24], and 1 (25%) [21] had the lowest QR of 5.

Clinical Effectiveness

The Wehrmann and Müller [24] pilot non-RCT used webcam footage to create a video-activity score to measure physical activity as an objective assessment of hyperactivity in children with suspected ADHD. The video-activity score did not show criterion validity with clinicians' or parents' hyperactivity ratings.

The findings from the ASD studies were more favorable. Reese and colleagues [22], who reported preliminary RCT findings on which families were assigned to in-clinic or telehealth evaluations, found that families could be coached to complete ASD assessment activities with young children via videoconferencing and clinicians could make accurate diagnoses remotely. Similarly, Juarez et al [21] reported on 2 studies, of which 1 compared a telediagnosis to a face-to-face assessment. This study demonstrated that, compared to gold-standard tools, remote ASD diagnostic consultations resulted in clinicians correctly diagnosing 78.9% (15/19) of children. No children were inaccurately diagnosed with ASD. Stainbrook and colleagues [23] investigated referrals before and after the introduction of a telehealth service. They found that implementing a diagnostic consultation service for ASD, in partnership with an early intervention service, increased referrals for diagnostic evaluation and the likelihood of families attending appointments. Following referral, 56 (89%) of the 63 families chose to receive further appointments via telehealth services rather than face-to-face services, and families with complex problems were the most likely to access clinic services.

User Impact, Feasibility, and Acceptability

In a second qualitative feasibility study, Juarez and colleagues [21] reported positive user feedback from both health care professionals and families. Families from rural areas reported geographical and time barriers to accessing traditional health care. These barriers were reduced with remote diagnoses, leading to high levels of satisfaction. Stainbrook and colleagues [23] found that families were more likely to attend telehealth appointments. Following referral, 56 (89%) of the 63 families chose to receive appointments via telehealth services rather than face-to-face services, and families with complex problems were the most likely to access clinic services.

Service Delivery Efficiencies and Economic Impact

Stainbrook and colleagues [23] were the only authors to document service delivery efficiencies. They reported that implementing a telehealth service reduced the time to diagnosis by 11-12 months.

Readiness for Clinic Adoption

Despite the effectiveness and positive user impact of telehealth in the assessment of ASD, the studies all had a small sample size. Of the 3 ASD papers, 1 (33%) did not report on suitability for implementation [23] and 2 (67%) stated that further research is necessary [21,22]. The ADHD assessment paper [24] reported negative findings and concluded that telehealth assessments for ADHD were not suitable for implementation. As such, prior to the COVID-19 pandemic, there were no telehealth technologies that were reported to be efficacious in assessing NDDs and suitable for immediate adoption in practice.

Monitoring

Summary of Monitoring Papers

We found 5 articles that reported on a sample of children with ADHD and their families. All papers were based on the CATTS (Children's ADHD Telemental Health Treatment Study) [25]. We found an RCT that assessed the effectiveness of a telehealth service for children with ADHD, which included pharmacological treatment monitoring and caregiver behavior training/psychoeducation. The RCT study compared families who received augmented treatment as usual, which involved only 1 telehealth consultation, to families who received 6 telehealth sessions, which were conducted approximately 1 month apart. The papers were generally highly rated (QR=2), and the main study was an RCT. Secondary papers looked at caregiver outcomes [26,27] and health care professionals' decisions on medication changes [28]. The remaining paper received a low QR (QR=5) because of the qualitative nature of the report, which focused on caregiver satisfaction and engagement, and health care professional fidelity [29]. Although not all studies were directly related to monitoring, they were collated together to allow the reader to understand that data were from multiple articles that related to the same trial. We found 4 studies that were based on the main trial's dataset, which included 223 families of children with ADHD aged 5-12 years and their carers [25]. The remaining study [27] involved a subsample of 37 participants.

Clinical Effectiveness

Overall, both methods of telehealth delivery resulted in reductions in ADHD and oppositional defiant disorder behaviors and improvements in role performance and impairment, with the telehealth model generally resulting in better outcomes [25] and better parental mental health [26] than face-to-face models.

In a feasibility trial of a subsample of families, Tse and colleagues [27] assessed the outcomes from baseline to 25 weeks and found similar outcomes for child ADHD behaviors. However, they also found that parents who used telehealth services had considerably less improvement in caregiver strain and empowerment than those who received face-to-face training.

User Impact, Feasibility, and Acceptability

High levels of engagement and satisfaction were reported by parents [27,29]. Rockhill et al [28] reported that fidelity was not impacted by telehealth delivery. The authors argued that telehealth provides added value in terms of increasing treat-to-target goals and offering support to health care professionals.

Service Delivery Efficiencies and Economic Impact

Service delivery efficiencies and economic impact were not reported in any monitoring papers.

Readiness for Clinic Adoption

Myers and colleagues [25] provided clinical guidance regarding the use of telehealth in treating children with ADHD, and the results from the RCT were promising. Further research is recommended in the development of the CATTS trial, including research that involves the greater use of teachers in interventions and objective school outcome measures, such as the completion of homework and behavioral observations, to further validate the tool. Tse and colleagues [27] concluded that telehealth delivery was promising in terms of readiness for clinic adoption, but telehealth for caregivers' distress needed further study, including the investigation of the best delivery modality. Future research on the cost benefits of telehealth models of care for ADHD was also recommended. These findings indicated promise in the implementation of technologies for monitoring ADHD.

Treatment

We found 33 papers that reported on the use of telehealth technologies to treat NDDs. The majority of the papers focused on ASD (18/33, 55%). Other reported conditions were ADHD with or without ASD (5/33, 15%), communication disorders (7/33, 21%), tic disorders (2/33, 6%) and learning disabilities (1/33, 3%). Due to the volume of treatment papers, each condition will be considered in turn.

ASD

Summary of ASD Treatment Papers

In terms of ASD, 1 paper presented a case report of a 16-year-old male with Asperger syndrome, social isolation, and depression. Clarke [30] reported that communicating via telehealth allowed a clinician to develop a relationship with a young person who was later able to attend a clinic in person and reconnect with his family. The remaining papers (17/18, 94%) focused on some aspect of parent training. Of these 17 papers, 6 (35%) reported on providing telehealth-delivered functional analysis and communication training to parents [31-36], and 1 (6%) reported on using telehealth-delivered functional analysis to train a health care professional [37]. These studies mainly consisted of case studies or case series (6/7, 86%) that used a multiple baseline experimental design and had a QR of 3 [35] or 4 [31,32,34,36,37]. Another paper (1/7, 14%) reexamined 2 nonresponding participants' data from an RCT (QR=4) [33].

We found 4 studies on 4 programs that incorporated self-directed online learning with remote therapy, support, or coaching [38-41]. These included studies on ImPACT Online Communication Training [38,39], which evaluated the feasibility (QR=4) and clinical efficacy (QR=3) of the ImPACT Online program in addressing social communication development, and a noncomparative feasibility study (QR=4) on OASIS ABA (Online and Applied System for Intervention Skills Applied Behavior Analysis)–based parent training [40]. Another program, which involved reciprocal imitation training, was used in a single-subject multiple-baseline design study (QR=4) [41].

In a noncomparative trial that gathered data from before and after intervention (QR=4), Little and colleagues [42] studied occupational-based coaching via telehealth for increasing positive interactions and everyday routines. This included an evaluation of acceptability/cost [43] and a linked qualitative (QR=5) appraisal of parents' perceptions [44]. The remaining programs were the Sunny Starts parent training program for increasing sociocommunicative behavior, which was used in a case series with multiple baseline experimental data (QR=4) [45]; the RUBI-PT (Research Unit on Behavioral Interventions-Parent Training) program, which was developed by the Research Unit on Behavioral Interventions Autism Network and targeted behaviors such as aggression and tantrums in children with ASD; benchmarking, which was used in a trial that compared the data of new services to data from previously published clinical trials (QR= 4) to evaluate effectiveness (eg, reduction in disruptive behavior), feasibility, and acceptability [46]; and parent coaching with a focus on educating parents about effective approaches for children with ASD (eg, social narratives and visual schedules), which was used in a qualitative paper (QR=5) [47].

Clinical Effectiveness

The majority of the ASD papers reported that treatment was clinically effective in improving caregiver knowledge, caregiver competence, and child participation (6/18, 33%) [38-42,45], increasing communication responses (2/18, 11%) [34,45], and reducing problem behaviors (5/18, 28%) [31,32,35-37]. We found 1 (5%) paper [33] that discussed 2 young children with ASD who underwent functional communication training, but this was unsuccessful in reducing problem behaviors. The authors suggested that although not all patients can be treated via telehealth, if sessions are recorded, watching the recordings can lead to the identification of the reason why treatment was not successful. Ingersoll and colleagues [39] noted that both online self-directed training and therapist-assisted, parent-mediated telehealth intervention led to improvements in fidelity, self-efficacy, stress, and parents' perceptions of their child, and that families who received therapist coaching and support gained improved social skills.

User Impact, Feasibility, and Acceptability

High levels of engagement and satisfaction were reported by parents [40,43,46]. However, difficulties surrounding failing technology and incomplete personal interaction were also documented [44,47]. Ingersoll and colleagues [38] found that parental engagement and satisfaction were similar for both self-directed and therapist-assisted methods of telehealth delivery. However, having a therapist increased engagement and led to higher rates of telehealth service completion. Parents often engaged with the program (ie, without therapist support) outside of traditional working hours, which allowed for greater flexibility than in face-to-face coaching.

In a qualitative study, Ashburner et al [47] explored the perceived advantages and disadvantages of a follow-up early intervention service that was delivered via remote technology, by comparing the service to previous face-to-face services. Content analysis showed that parents, service providers, and the ASD specialist perceived remote technologies to be helpful in upskilling parents/service providers and enabling families to access support from home. However, all study participants agreed that remote technology should be used to augment, rather than replace, face-to-face contact, which is similar to the findings reported by Little et al [43].

The use of telehealth for training health care professionals was also shown to be a promising way of providing support to practitioners in the field, which led to the greater implementation of target strategies [37].

Service Delivery Efficiencies and Economic Impact

Suess and colleagues [32] reported on telehealth service delivery efficiencies and argued that in some cases, brief, efficient telehealth appointments bypasses the need for further in-clinic support and allows for quicker treatment initiation. Several authors [40,41,47] suggested that telehealth has the potential to increase access to ASD services (ie, particularly in remote areas) and reduce costs, time, and travel. Lingren et al [35] compared the costs of therapy for caregivers of children with ASD between different telehealth models, including in-home telehealth, regional clinic telehealth, and in-home, face-to-face telehealth models. The costs were lowest for the in-home telehealth model, but the in-home and in-home, face-to-face telehealth models were substantially less costly than the costs for face-to-face in-home therapy. Similarly, in a study that involved a 12-week telehealth intervention for families with a child with ASD, the authors reported that the costs for both outpatient and in-home care models were approximately 2.6 times more expensive than the costs for telehealth models [43].

ADHD

Summary of ADHD Treatment Papers

We found 5 treatment papers from the ADHD sample that included patients with ADHD and patients with ASD. The highest quality paper (ie, a small RCT with a QR of 2) compared patients who underwent internet-based cognitive behavioral therapy based on the InFocus program (ie, with/without therapist support) to those in the waitlist control [48]. The other papers were of much lower in quality. We found 1 experimental pilot study (QR=5) that used a nonrandomized pre-post intervention study design to assess the feasibility and acceptability of a parenting group training program delivered via telehealth [49]. Additionally, Sehlin and colleagues [50] provided qualitative data (QR=5) for a study that involved a face-to-face meeting that was followed by 8 weeks of internet-based chat sessions for providing coaching and support. Another qualitative paper (QR=5) conducted implementation interviews with health care professionals after providing coaching and support at 3 trial sites in Sweden [51]. The final paper used a multiple descriptive case design (QR=4) to assess caregiver perspectives in a sample of 10 caregivers of young people with ADHD or ASD who took part in an internet-based intervention [52].

Clinical Effectiveness

In general, clinical effectiveness was unclear or not reported (3/5, 60%) [50-52], and all studies were limited by small sample sizes (range 7-45). We found 2 group therapy telehealth programs that showed great promise. A study [48] found that an internet-based cognitive behavioral therapy telehealth treatment program, which included weekly online group therapy sessions for adults with ADHD, was no more clinically effective than unsupported self-help alone. However, people in both programs faired better than those in the waiting list controls. The second group therapy program showed a trend of improvement in child ADHD symptoms following a group parenting intervention, but the program was not adequately powered [49].

User Impact, Feasibility, and Acceptability

Sehlin et al [50] found that although remote coaching was perceived favorably by participants, difficulties surrounded failing technology and incomplete personal interaction were reported. Shah and colleagues [49] also reported that clinicians experienced difficulties with internet connections and found it hard to read body language and expressions, as faces were sometimes out of focus during video appointments. They also reported that patients experienced disturbances from other family members, and that the inability to role play during telehealth appointments was problematic. However, parents were at ease and relaxed during telehealth appointments.

Gillberg and Wentz [51] assessed professionals' perceptions on internet-based support and coaching and the barriers and facilitators to implementation. Facilitators of positive perceptions included improved access, equality distribution, and the delivery/quality of health care services. Reported barriers included the design of the intervention, technical issues, attitudes of staff, organizational culture and structure, and work division and resource allocation.

Service Delivery Efficiencies and Economic Impact

Most studies (4/5, 80%) did not report the economic impact. However, cost-savings in terms of time and travel were noted in 1 (20%) study [49].

Other NNDs (ie, Communication Disorders, Tic Disorders, and Learning Disabilities)

Summary of Other Treatment Papers

Treatment programs for communication disorders (eg, stuttering) were evaluated in 7 papers. A noncomparative trial (QR=4) investigated the Camperdown Program, which was used to reduce stuttering in adolescents [53]. The remaining papers assessed the Lidcombe program for preschoolers. With regard to the Lidcombe program, we found 1 RCT (QR=2) that compared telehealth care delivery to in-clinic, face-to-face care delivery [54], and 1 study (QR=2) that involved a quantitative evaluation of parent satisfaction ratings [55]. We found 1 paper (QR=4) that involved a noncomparative trial that assessed reductions in stuttering severity and frequency, as well as satisfaction with telehealth delivery [56]. The remaining papers all had the lowest quality rating (QR=5). We found 2 papers that discussed clinical insights from health care professionals who were involved in telehealth treatment delivery for patients with stuttering [57,58]. We also found a descriptive-analytic study of satisfaction with telehealth treatment for stuttering [59]. Furthermore, we found 2 papers on a pilot open-case series (QR=4) [60] and an RCT that used a waiting list control for the assessment and treatment of chronic tic disorders (QR=2) [61]. The final paper (QR=5) provided an account of a telehealth service that was delivered at a large-scale regional service level [62].

Clinical Effectiveness

In terms of the Lidcombe program, the Phase I [56] trial demonstrated the efficacy of remotely delivering the program to families with a preschool child who stutters. However, the results of a main parallel, open-plan, noninferiority RCT trial [54] showed that it was not clear whether webcam treatment was noninferior to standard treatment in the short term. Carey and colleagues [53] conducted a Phase II clinical trial that examined adolescents' responsiveness to the webcam-delivered Camperdown program, and found that adolescents experienced substantially reduced stuttering in terms of both frequency and severity, although relapse was a problem.

User Impact, Feasibility, and Acceptability

High levels of engagement and satisfaction were reported by parents/carers [53,60,61]. The use of telehealth to train health care professionals was shown to be a promising way of providing support to practitioners in the field [58]. Jahromi and Ahmadian [59] explored satisfaction in telespeech therapy among 30 Iranian patients aged ≥14 years. The authors reported that satisfaction with the therapy was high, but the low internet speed in the country was a major challenge for half the participants, as they could not maintain eye contact with the therapist due to the distorted image transmission. Similarly, another study reported that difficulties arose with regard to completing certain aspects of treatment due to limited web camera viewing ranges and audio/visual difficulties [61].

The feasibility of delivering both the Lidcombe and Camperdown programs via telehealth methods was documented, and parents were generally satisfied [53,54,56]. However, Bridgman and colleagues [57] highlighted that individual adjustments were required to tailor the treatment process to families' needs in order to maximize outcomes. Ferdinands and Bridgman [55] examined parent satisfaction and stuttering severity at baseline and during the 9-month/18-month follow up, and found that increased parental satisfaction was generally, but not always, linked with the severity of stuttering. This demonstrates the need to provide treatment at the family level when monitoring children with communication disorders. There was no considerable difference in parent satisfaction between clinic and telehealth care delivery.

Ricketts et al [60,61] conducted pilot studies that explored the feasibility of assessing tic severity over voice over internet protocol (VoIP), which allows users to make and receive calls via an internet connection. They compared the feasibility, acceptability, and efficacy of VoIP-delivered therapy for tic disorders to those of a waitlist control. They found a decrease in tic severity that was similar to the decrease identified in the original Cognitive Behavioral Intervention for Tics trial [63] and greater than that of the waitlist control [61].

Service Delivery Efficiencies and Economic Impact

Merrill et al [62] provided an overview of Ohio's Telepsychiatry Project for Intellectual Disability, which provides specialized mental health services to rural communities. This paper documented telehealth from a service delivery perspective. Although no specific figures were given, the report indicated that the service improved access to care, reduced emergency department visits/hospitalizations, and resulted in cost savings, including reduced travel expenses, medical expenses, and support costs. Similar cost savings were reported in other studies [54,59].

Readiness for Clinic Adoption

Of the 33 treatment papers, 5 (15%) deemed telehealth to be suitable for clinic adoption, either as an adjunct to current practices or on its own [30,47,50,59,62]. Furthermore, 24 (73%) papers noted that telehealth required further research before being implemented into clinical practice. The remaining papers were unclear/did not report on readiness for clinic adoption. Prior to the COVID-19 pandemic, the delivery of interventions via telehealth for parents of children with ASD, young people and adults who stutter, and adults with intellectual disability were thought to be suitable for clinic adoption. For young people who struggle with attending appointments, therapy conducted via VoIP was recommended. In addition, coaching and support via a chat program was recommended as an adjunct to usual treatment for young people and adults with ADHD/ASD.

Discussion

Principal Results

The purpose of this systematic review was to examine the evidence base for the clinical use of technology within the neurodevelopmental field prior to the COVID-19 pandemic, to identify possible telehealth technologies that can be considered for wide-spread implementation and document the current state of the evidence base prior to the anticipated rapid development in this field.

Assessment

Telehealth has been used to assess small samples of people with ASD, and telehealth shows promise for clinical adoption. In terms of economic impact, there are potential cost savings and service efficiencies, but the evidence base is limited. The ADHD assessment tool is not clinically effective, and there has been no evidence for the assessment of other NDDs at present.

Monitoring

As identified in a previous review by Spencer and colleagues [15], all studies that used telehealth for monitoring were for monitoring ADHD, and all studies were from the same trial (ie, the CATTS [25]), which had promising results for acceptability and effectiveness. Telehealth monitoring seems to be an approach that should be considered for clinic adoption.

Treatment

Telehealth has been used to treat a range of NDDs. However, the majority of conditions fall under ASD, and treatment has mainly focused on parent training interventions. These interventions have shown some evidence of clinical efficacy, such as improving caregiver knowledge, competence, and child participation, and reducing problem behaviors. Even when telehealth is not clinically effective, the recording of sessions can help health care professionals identify why the treatment did not work [33]. Treatments for communication disorders have also focused on parent intervention programs, which have shown some evidence of clinical efficacy and no difference in parent satisfaction between remote delivery and face-to-face delivery. Despite the fact that previous literature has suggested that the evidence for using telehealth to manage communication disorders is substantial [10], our review did not reveal a large number of papers that involved communication disorders, as more papers focused on ASD. Furthermore, our findings on effectiveness were mixed; the lack of an adequate control group was a limiting factor in several studies [53].

We found little evidence for the delivery of parenting interventions for ADHD. However, it is possible that the search terms used in this review limited access to such papers. Telehealth services for young people and adult service users tended to focus on the remote delivery of coaching, support, and therapy. The 1 case study of a young person with ASD who received online therapy had a promising outcome. People with ASD may particularly benefit from using technology to overcome communication difficulties, as this involves fewer social pressures than face-to-face therapy [64].

Cognitive-behavioral strategies have been used for both ADHD and tic disorders, and mixed clinical efficacies have been reported. There is a larger body of evidence for using behavioral and cognitive-behavioral treatments for tic disorders than evidence for using such treatments for ADHD, but further research is necessary for both disorders. There was limited evidence for using telehealth as a means of providing training to health care professionals. However, barriers to this approach, including the design of the intervention, technical issues, attitudes of staff, organizational culture and structure, and work division and resource allocation, were widely reported.

In summary, there is a much larger body of evidence for the efficacy of providing remotely delivered interventions to parents and children than evidence for providing such interventions to young people and adult service users. There is also a small body of evidence for using telehealth to train health care professionals. Generally, the user impact for all participants was positive. There was very little research on economic impact. Overall, the evidence base is of variable quality.

Key Implementation Issues

This systematic review highlighted key implementation issues for using telehealth services. The number of telehealth technologies that are ready to be implemented in practice is limited, as most studies stated that further research is necessary before such technologies are acceptable for clinical adoption. Service providers should consider both service users' opinions on such technology and the evidence base when choosing whether to implement telehealth technology into clinical practice. If families view telehealth technology as an adjunct to usual treatment, cost savings may not be achieved.

The telehealth delivery of treatments may have benefits. In some studies, allowing users to access treatment at convenient times and providing personalized treatment led to greater treatment engagement and completion. Families were more at ease and relaxed when participating in telehealth treatment. The need to personalize treatment to individuals and families was apparent across several studies. This is particularly important, as disruptions by other family members can occur. There is limited evidence for service delivery efficiencies. Implementation difficulties included failing technology, audio and visual problems, and difficulties in making eye contact. These were particularly problematic in countries with low internet speeds. In line with previous reviews, several studies have reported that health care professionals found reading body language and facial expressions difficult due to distorted images [10,65].

Directions for Future Research

This systematic review reveals that there is a lack of research that assesses the use of telehealth in aiding the diagnosis of a wide range of NDDs, and that the current focus is on autism. In general, cost-effectiveness and possible service efficiencies are underinvestigated, but they are an important consideration for real-world implementation. Future research should focus on developing guidelines and blueprints for how to best integrate telehealth care into clinical practice [66].

Limitations

The limitations of this study must be taken into account when interpreting the findings. As the search yielded a much greater number of papers than anticipated, the search was limited to the previous 5 years. This was a deviation from the initial protocol. However, it can be argued that this method allows for a more effective analysis of current technology and precludes the inclusion of outdated technology. Furthermore, limiting the search to published academic papers may have exacerbated the risk of bias, as authors were not contacted for unpublished work due to the volume of published papers obtained. This is a limitation of our study, and further reviews should explore unpublished data, especially data from conference papers, as these provided a vast amount of possibly relevant data. However, conference papers were excluded from this systematic review due to time constraints.

The majority of studies were conducted in high-income countries, thereby limiting the generalizability of our findings. It is likely that there would be intercountry variations in barriers to implementing new technology into existing health care systems. Although these barriers are typically considered outside the remit of standard reporting for trials, an understanding of these barriers is important if these technologies are to be routinely implemented.

The majority of data were of mid- to low-quality, and our findings should be interpreted with caution. This was generally because of small sample sizes and the high number of qualitative/reflexive study designs. However, RCTs are time-consuming and do not always lend themselves to real-world evaluations.

Conclusions

Our literature search highlighted that, prior to the COVID-19 pandemic, there was promising evidence for the use of telehealth in clinical practice, in relation to NDDs. Telehealth technologies were more frequently used to support the treatment and monitoring of NDDs; there was less evidence for their use in supporting the assessment of NDDs. The main focus of telehealth in the neurodevelopmental field was on ASD and ADHD, which are two of the most commonly occurring NDDs. There was evidence of good clinical outcomes and cost savings for health care providers. However, further research is required to substantiate this evidence. With the growing need to provide easy access to remotely delivered clinical support for enabling the wide-spread reach of health care and reducing the risk of spreading infectious diseases, it is essential that real-world evaluations for implementation and cost-effectiveness are conducted.

Acknowledgments

We would like to thank Nikki Brown, Jen Martin, Bethan Davies, Emma Rowley, Julie Clarke and, Joe Kilgariff for their support in conducting this systematic review.

This research was funded by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care East Midlands. The funders had no role in the study design; the collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication.

The views expressed are those of the author(s) and not necessarily those of the United Kingdom National Health Service, the National Institute for Health Research, or the Department of Health.

Authors' Contributions

CLH designed the study and wrote the protocol; MJG and CH provided feedback on the protocol. EY and AZV conducted the literature searches. AZV, CLH, BJB, and SSH reviewed the abstracts and titles for inclusion. AZV and CLH extracted the data. CLH and BJB completed the risk of bias assessment. AZV wrote the first draft of the manuscript with support from CLH and SSH. CLH supervised the process with support from CH. All authors contributed to and approved the final manuscript

Conflicts of Interest

CH acknowledges the financial support of the National Institute of Health Research Nottingham Biomedical Research Centre and National Institute of Health Research MindTech MedTech Co-operative. All other authors declare that they have no conflicts of interest.

Multimedia Appendix 1
Sample search for 1 database (ie, Medline).
[DOC File , 44 KB - jmir_v23i1e22619_app1.doc]

Multimedia Appendix 2
Summary of included studies.
[DOC File , 139 KB - jmir_v23i1e22619_app2.doc]

References

1. Freeman RD, Fast DK, Burd L, Kerbeshian J, Robertson MM, Sandor P. An international perspective on Tourette syndrome: selected findings from 3,500 individuals in 22 countries. Dev Med Child Neurol 2000 Jul;42(7):436-447 [FREE Full text] [doi: 10.1017/s0012162200000839] [Medline: 10972415]
2. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry 2008 Aug;47(8):921-929. [doi: 10.1097/CHI.0b013e318179964f] [Medline: 18645422]
3. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychological Association; 2013.
4. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
5. Speciality guides for patient management during the coronavirus pandemic: Clinical guide for frontline staff to support the management of patients with a learning disability, autism or both during the coronavirus pandemic – relevant to all clinical

specialities. National Health Service England and National Health Service Improvement. 2020 Nov. URL: https://www.nice.org.uk/Media/Default/About/COVID-19/Specialty-guides/learning-disability-autism-during-pandemic.pdf [accessed 2021-01-04]
6. Kola L. Global mental health and COVID-19. Lancet Psychiatry 2020 Aug;7(8):655-657 [FREE Full text] [doi: 10.1016/S2215-0366(20)30235-2] [Medline: 32502468]
7. Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
8. Pereira-Sanchez V, Adiukwu F, El Hayek S, Bytyçi CG, Gonzalez-Diaz JM, Kundadak GK, et al. COVID-19 effect on mental health: patients and workforce. Lancet Psychiatry 2020 Jun;7(6):e29-e30 [FREE Full text] [doi: 10.1016/S2215-0366(20)30153-X] [Medline: 32445691]
9. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
10. Molini-Avejonas DR, Rondon-Melo S, de La Higuera Amato CA, Samelli AG. A systematic review of the use of telehealth in speech, language and hearing sciences. J Telemed Telecare 2015 Oct;21(7):367-376. [doi: 10.1177/1357633X15583215] [Medline: 26026181]
11. Lavorgna L, Brigo F, Abbadessa G, Bucello S, Clerico M, Cocco E, et al. The Use of Social Media and Digital Devices Among Italian Neurologists. Front Neurol 2020;11:583 [FREE Full text] [doi: 10.3389/fneur.2020.00583] [Medline: 32612572]
12. Sutherland R, Trembath D, Roberts J. Telehealth and autism: A systematic search and review of the literature. Int J Speech Lang Pathol 2018 Jun;20(3):324-336. [doi: 10.1080/17549507.2018.1465123] [Medline: 29709201]
13. Boisvert M, Lang R, Andrianopoulos M, Boscardin ML. Telepractice in the assessment and treatment of individuals with autism spectrum disorders: A systematic review. Dev Neurorehabil 2010;13(6):423-432. [doi: 10.3109/17518423.2010.499889] [Medline: 20887200]
14. Aresti-Bartolome N, Garcia-Zapirain B. Technologies as support tools for persons with autistic spectrum disorder: a systematic review. Int J Environ Res Public Health 2014 Aug 04;11(8):7767-7802 [FREE Full text] [doi: 10.3390/ijerph110807767] [Medline: 25093654]
15. Spencer T, Noyes E, Biederman J. Telemedicine in the Management of ADHD: Literature Review of Telemedicine in ADHD. J Atten Disord 2020 Jan;24(1):3-9. [doi: 10.1177/1087054719859081] [Medline: 31257978]
16. Fagherazzi G, Goetzinger C, Rashid M, Aguayo G, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res. Preprint posted online on April 11, 2020. [doi: 10.2196/preprints.19284]
17. Valentine AZ, Brown BJ, Groom MJ, Young E, Hollis C, Hall CL. A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: A map of the current evidence. Clin Psychol Rev 2020 Aug;80:101870 [FREE Full text] [doi: 10.1016/j.cpr.2020.101870] [Medline: 32712216]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
19. Timeline: WHO's COVID-19 response. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline [accessed 2021-01-04]
20. OCEBM Levels of Evidence. Oxford Centre for Evidence-Based Medicine. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence [accessed 2021-01-04]
21. Juárez AP, Weitlauf AS, Nicholson A, Pasternak A, Broderick N, Hine J, et al. Early Identification of ASD Through Telemedicine: Potential Value for Underserved Populations. J Autism Dev Disord 2018 Aug;48(8):2601-2610 [FREE Full text] [doi: 10.1007/s10803-018-3524-y] [Medline: 29527626]
22. Reese RM, Jamison TR, Braun M, Wendland M, Black W, Hadorn M, et al. Brief report: use of interactive television in identifying autism in young children: methodology and preliminary data. J Autism Dev Disord 2015 May;45(5):1474-1482 [FREE Full text] [doi: 10.1007/s10803-014-2269-5] [Medline: 25331323]
23. Stainbrook JA, Weitlauf AS, Juárez AP, Taylor JL, Hine J, Broderick N, et al. Measuring the service system impact of a novel telediagnostic service program for young children with autism spectrum disorder. Autism 2019 May;23(4):1051-1056. [doi: 10.1177/1362361318787797] [Medline: 30112912]
24. Wehrmann T, Müller JM. An objective measure of hyperactivity aspects with compressed webcam video. Child Adolesc Psychiatry Ment Health 2015;9:45 [FREE Full text] [doi: 10.1186/s13034-015-0076-1] [Medline: 26361496]
25. Myers K, Vander Stoep A, Zhou C, McCarty CA, Katon W. Effectiveness of a telehealth service delivery model for treating attention-deficit/hyperactivity disorder: a community-based randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2015 Apr;54(4):263-274 [FREE Full text] [doi: 10.1016/j.jaac.2015.01.009] [Medline: 25791143]

26. Vander Stoep A, McCarty CA, Zhou C, Rockhill CM, Schoenfelder EN, Myers K. The Children's Attention-Deficit Hyperactivity Disorder Telemental Health Treatment Study: Caregiver Outcomes. J Abnorm Child Psychol 2017 Jan;45(1):27-43 [FREE Full text] [doi: 10.1007/s10802-016-0155-7] [Medline: 27117555]
27. Tse YJ, McCarty CA, Stoep AV, Myers KM. Teletherapy delivery of caregiver behavior training for children with attention-deficit hyperactivity disorder. Telemed J E Health 2015 Jun;21(6):451-458. [doi: 10.1089/tmj.2014.0132] [Medline: 25719609]
28. Rockhill CM, Tse YJ, Fesinmeyer MD, Garcia J, Myers K. Telepsychiatrists' Medication Treatment Strategies in the Children's Attention-Deficit/Hyperactivity Disorder Telemental Health Treatment Study. J Child Adolesc Psychopharmacol 2016 Oct;26(8):662-671 [FREE Full text] [doi: 10.1089/cap.2015.0017] [Medline: 26258927]
29. McCarty CA, Vander Stoep A, Violette H, Myers K. Interventions Developed for Psychiatric and Behavioral Treatment in the Children’s ADHD Telemental Health Treatment Study. J Child Fam Stud 2014 May 17;24:1735-1743 [FREE Full text] [doi: 10.1007/s10826-014-9977-5]
30. Clarke CS. Telepsychiatry in Asperger's syndrome. Ir J Psychol Med 2018 Dec;35(4):325-328. [doi: 10.1017/ipm.2017.19] [Medline: 30501669]
31. Suess AN, Romani PW, Wacker DP, Dyson SM, Kuhle JL, Lee JF, et al. Evaluating the Treatment Fidelity of Parents Who Conduct In-Home Functional Communication Training with Coaching via Telehealth. J Behav Educ 2013 Sep 14;23:34-59. [doi: 10.1007/s10864-013-9183-3]
32. Suess AN, Wacker DP, Schwartz JE, Lustig N, Detrick J. Preliminary evidence on the use of telehealth in an outpatient behavior clinic. J Appl Behav Anal 2016 Sep;49(3):686-692. [doi: 10.1002/jaba.305] [Medline: 27001117]
33. Schieltz KM, Romani PW, Wacker DP, Suess AN, Huang P, Berg WK, et al. Single-Case Analysis to Determine Reasons for Failure of Behavioral Treatment via Telehealth. Remedial Spec Educ 2018 Jan 07;39(2):95-105 [FREE Full text] [doi: 10.1177/0741932517743791]
34. Simacek J, Dimian AF, McComas JJ. Communication Intervention for Young Children with Severe Neurodevelopmental Disabilities Via Telehealth. J Autism Dev Disord 2017 Mar;47(3):744-767 [FREE Full text] [doi: 10.1007/s10803-016-3006-z] [Medline: 28093677]
35. Lindgren S, Wacker D, Suess A, Schieltz K, Pelzel K, Kopelman T, et al. Telehealth and Autism: Treating Challenging Behavior at Lower Cost. Pediatrics 2016 Feb;137 Suppl 2(Suppl 2):S167-S175 [FREE Full text] [doi: 10.1542/peds.2015-2851O] [Medline: 26908472]
36. Suess AN, Schieltz KM, Wacker DP, Detrick J, Podlesnik CA. An evaluation of resurgence following functional communication training conducted in alternative antecedent contexts via telehealth. J Exp Anal Behav 2020 Jan;113(1):278-301. [doi: 10.1002/jeab.551] [Medline: 31617951]
37. Fettig A, Barton EE, Carter AS, Eisenhower AS. Using e-coaching to support an early intervention provider's implementation of a functional assessment-based intervention. Infants Young Child 2016;29(2):130-147 [FREE Full text] [doi: 10.1097/IYC.0000000000000058] [Medline: 30337773]
38. Ingersoll B, Berger NI. Parent Engagement With a Telehealth-Based Parent-Mediated Intervention Program for Children With Autism Spectrum Disorders: Predictors of Program Use and Parent Outcomes. J Med Internet Res 2015 Oct 06;17(10):e227 [FREE Full text] [doi: 10.2196/jmir.4913] [Medline: 26443557]
39. Ingersoll B, Wainer AL, Berger NI, Pickard KE, Bonter N. Comparison of a Self-Directed and Therapist-Assisted Telehealth Parent-Mediated Intervention for Children with ASD: A Pilot RCT. J Autism Dev Disord 2016 Feb 27;46(7):2275-2284. [doi: 10.1007/s10803-016-2755-z]
40. Heitzman-Powell LS, Buzhardt J, Rusinko LC, Miller TM. Formative Evaluation of an ABA Outreach Training Program for Parents of Children With Autism in Remote Areas. Focus Autism Other Dev Disabl 2013 Oct 14;29(1):23-38. [doi: 10.1177/1088357613504992]
41. Wainer AL, Ingersoll BR. Increasing Access to an ASD Imitation Intervention Via a Telehealth Parent Training Program. J Autism Dev Disord 2015 Dec;45(12):3877-3890. [doi: 10.1007/s10803-014-2186-7] [Medline: 25035089]
42. Little LM, Pope E, Wallisch A, Dunn W. Occupation-Based Coaching by Means of Telehealth for Families of Young Children With Autism Spectrum Disorder. Am J Occup Ther 2018;72(2):7202205020p1-7202205020p7. [doi: 10.5014/ajot.2018.024786] [Medline: 29426380]
43. Little LM, Wallisch A, Pope E, Dunn W. Acceptability and Cost Comparison of a Telehealth Intervention for Families of Children With Autism. Infants Young Child 2018;31(4):275-286. [doi: 10.1097/iyc.0000000000000126]
44. Wallisch A, Little L, Pope E, Dunn W. Parent Perspectives of an Occupational Therapy Telehealth Intervention. Int J Telerehabil 2019;11(1):15-22 [FREE Full text] [doi: 10.5195/ijt.2019.6274] [Medline: 31341543]
45. Guðmundsdóttir K, Sigurðardóttir ZG, Ala'i-Rosales S. Evaluation of Caregiver Training via Telecommunication for Rural Icelandic Children With Autism. Behav Dev Bull 2017 Apr;22(1):215-229. [doi: 10.1037/bdb0000040]
46. Bearss K, Burrell TL, Challa SA, Postorino V, Gillespie SE, Crooks C, et al. Feasibility of Parent Training via Telehealth for Children with Autism Spectrum Disorder and Disruptive Behavior: A Demonstration Pilot. J Autism Dev Disord 2018 Apr;48(4):1020-1030. [doi: 10.1007/s10803-017-3363-2] [Medline: 29170938]

47. Ashburner J, Vickerstaff S, Beetge J, Copley J. Remote versus face-to-face delivery of early intervention programs for children with autism spectrum disorders: Perceptions of rural families and service providers. Res Autism Spectr Disord 2016 Mar;23:1-14. [doi: 10.1016/j.rasd.2015.11.011]
48. Pettersson R, Söderström S, Edlund-Söderström K, Nilsson KW. Internet-Based Cognitive Behavioral Therapy for Adults With ADHD in Outpatient Psychiatric Care. J Atten Disord 2017 Apr;21(6):508-521. [doi: 10.1177/1087054714539998] [Medline: 24970720]
49. Shah R, Chakrabarti S, Sharma A, Grover S, Sachdeva D, Avasthi A. Participating from homes and offices: Proof-of-concept study of multi-point videoconferencing to deliver group parent training intervention for attention-deficit/ hyperactivity disorder. Asian J Psychiatr 2019 Mar;41:20-22. [doi: 10.1016/j.ajp.2019.03.006] [Medline: 30877843]
50. Sehlin H, Hedman Ahlström B, Andersson G, Wentz E. Experiences of an internet-based support and coaching model for adolescents and young adults with ADHD and autism spectrum disorder -a qualitative study. BMC Psychiatry 2018 Jan 18;18(1):15 [FREE Full text] [doi: 10.1186/s12888-018-1599-9] [Medline: 29347983]
51. Gillberg N, Wentz E. Internet-Based Support and Coaching for Adolescents and Young Adults with Neuropsychiatric Disorders—The Implementation of an Intervention from an Organizational Perspective. Health 2017 Jan 13;09(01):69-87. [doi: 10.4236/health.2017.91006]
52. Söderqvist H, Kajsa E, Ahlström BH, Wentz E. The caregivers' perspectives of burden before and after an internet-based intervention of young persons with ADHD or autism spectrum disorder. Scand J Occup Ther 2017 Sep;24(5):383-392. [doi: 10.1080/11038128.2016.1267258] [Medline: 28140737]
53. Carey B, O'Brian S, Lowe R, Onslow M. Webcam delivery of the Camperdown Program for adolescents who stutter: a phase II trial. Lang Speech Hear Serv Sch 2014 Oct;45(4):314-324. [doi: 10.1044/2014_LSHSS-13-0067] [Medline: 25091362]
54. Bridgman K, Onslow M, O'Brian S, Jones M, Block S. Lidcombe Program Webcam Treatment for Early Stuttering: A Randomized Controlled Trial. J Speech Lang Hear Res 2016 Oct 01;59(5):932-939. [doi: 10.1044/2016_JSLHR-S-15-0011] [Medline: 27617680]
55. Ferdinands B, Bridgman K. An investigation into the relationship between parent satisfaction and child fluency in the Lidcombe Program: Clinic versus telehealth delivery. Int J Speech Lang Pathol 2019 Aug;21(4):347-354. [doi: 10.1080/17549507.2018.1445779] [Medline: 29621894]
56. O'Brian S, Smith K, Onslow M. Webcam delivery of the Lidcombe program for early stuttering: a phase I clinical trial. J Speech Lang Hear Res 2014 Jun 01;57(3):825-830. [doi: 10.1044/2014_JSLHR-S-13-0094] [Medline: 24686834]
57. Bridgman K, Block S, O'Brian S. Webcam delivery of the Lidcombe Program: insights from a clinical trial. ACQuiring Knowledge in Speech, Language and Hearing 2015;17(3):125-129.
58. Bridgman K, Pallathil K, Ford N, Tran J, Lam D, Wee W, et al. Attitudes and experiences of SLP students from a pilot telehealth stuttering clinic. ACQuiring Knowledge in Speech, Language and Hearing 2018;20(1):14-20.
59. Jahromi ME, Ahmadian L. Evaluating satisfaction of patients with stutter regarding the tele-speech therapy method and infrastructure. Int J Med Inform 2018 Jul;115:128-133. [doi: 10.1016/j.ijmedinf.2018.03.004] [Medline: 29779715]
60. Ricketts EJ, Bauer CC, Ran D, Himle MB, Woods DW. Pilot Open Case Series of Voice over Internet Protocol-delivered Assessment and Behavior Therapy for Chronic Tic Disorders. Cogn Behav Pract 2016 Feb;23(1):40-50 [FREE Full text] [doi: 10.1016/j.cbpra.2014.09.003] [Medline: 30595642]
61. Ricketts EJ, Goetz AR, Capriotti MR, Bauer CC, Brei NG, Himle MB, et al. A randomized waitlist-controlled pilot trial of voice over Internet protocol-delivered behavior therapy for youth with chronic tic disorders. J Telemed Telecare 2016 Apr;22(3):153-162 [FREE Full text] [doi: 10.1177/1357633X15593192] [Medline: 26169350]
62. Merrill B, Cowan AE, Gentile JP. House Calls: Telepsychiatry with Patients with Intellectual Disability. Ann Med Health Sci Res 2017;7:463-465.
63. Piacentini J, Woods DW, Scahill L, Wilhelm S, Peterson AL, Chang S, et al. Behavior therapy for children with Tourette disorder: a randomized controlled trial. JAMA 2010 May 19;303(19):1929-1937 [FREE Full text] [doi: 10.1001/jama.2010.607] [Medline: 20483969]
64. Goodwin MS. Enhancing and Accelerating the Pace of Autism Research and Treatment: The Promise of Developing Innovative Technology. Focus Autism Other Dev Disabl 2008 Mar 26;23(2):125-128. [doi: 10.1177/1088357608316678]
65. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]
66. Balcombe L, De Leo D. An Integrated Blueprint for Digital Mental Health Services Amidst COVID-19. JMIR Ment Health 2020 Jul 22;7(7):e21718 [FREE Full text] [doi: 10.2196/21718] [Medline: 32668402]

Abbreviations

ADHD: attention deficit hyperactivity disorder
ASD: autism spectrum disorder
CATTS: Children’s ADHD Telemental Health Treatment Study
NDD: neurodevelopmental disorder

OASIS ABA: Online and Applied System for Intervention Skills Applied Behavior Analysis
PICOS: population, intervention, comparison, outcome, study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QR: quality rating
RCT: randomized controlled trial
RUBI-PT: Research Unit on Behavioral Interventions-Parent Training
VoIP: voice over internet protocol

Edited by G Eysenbach; submitted 18.07.20; peer-reviewed by L Le, L Lavorgna; comments to author 25.08.20; revised version received 29.09.20; accepted 26.10.20; published 20.01.21.

Please cite as:
Valentine AZ, Hall SS, Young E, Brown BJ, Groom MJ, Hollis C, Hall CL
Implementation of Telehealth Services to Assess, Monitor, and Treat Neurodevelopmental Disorders: Systematic Review
J Med Internet Res 2021;23(1):e22619
URL: *http://www.jmir.org/2021/1/e22619/*
doi:*10.2196/22619*
PMID:*33326409*

Review

Factors Affecting the Implementation, Use, and Adoption of Real-Time Location System Technology for Persons Living With Cognitive Disabilities in Long-term Care Homes: Systematic Review

Alisa Grigorovich[1], PhD; Yalinie Kulandaivelu[1,2], MSc; Kristine Newman[3], RN, CRN, PhD; Andria Bianchi[4,5], PhD; Shehroz S Khan[1], PhD; Andrea Iaboni[1,6], DPhil, MD; Josephine McMurray[7], MRT, MBA, PhD

[1]KITE - Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada
[2]Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[3]Daphne Cockwell School of Nursing, Ryerson University, Toronto, ON, Canada
[4]Bioethics Program, University Health Network, Toronto, ON, Canada
[5]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[6]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[7]Lazaridis School of Business & Economics, Wilfred Laurier University, Brantford, ON, Canada

Corresponding Author:
Josephine McMurray, MRT, MBA, PhD
Lazaridis School of Business & Economics
Wilfred Laurier University
SC Johnston Building, Room 304
73 George Street
Brantford, ON, N3T 3Y3
Canada
Phone: 1 15197568228 ext 5649
Email: jmcmurray@wlu.ca

Abstract

Background: As the aging population continues to grow, the number of adults living with dementia or other cognitive disabilities in residential long-term care homes is expected to increase. Technologies such as real-time locating systems (RTLS) are being investigated for their potential to improve the health and safety of residents and the quality of care and efficiency of long-term care facilities.

Objective: The aim of this study is to identify factors that affect the implementation, adoption, and use of RTLS for use with persons living with dementia or other cognitive disabilities in long-term care homes.

Methods: We conducted a systematic review of the peer-reviewed English language literature indexed in MEDLINE, Embase, PsycINFO, and CINAHL from inception up to and including May 5, 2020. Search strategies included keywords and subject headings related to cognitive disability, residential long-term care settings, and RTLS. Study characteristics, methodologies, and data were extracted and analyzed using constant comparative techniques.

Results: A total of 12 publications were included in the review. Most studies were conducted in the Netherlands (7/12, 58%) and used a descriptive qualitative study design. We identified 3 themes from our analysis of the studies: barriers to implementation, enablers of implementation, and agency and context. Barriers to implementation included lack of motivation for engagement; technology ecosystem and infrastructure challenges; and myths, stories, and shared understanding. Enablers of implementation included understanding local workflows, policies, and technologies; usability and user-centered design; communication with providers; and establishing policies, frameworks, governance, and evaluation. Agency and context were examined from the perspective of residents, family members, care providers, and the long-term care organizations.

Conclusions: There is a striking lack of evidence to justify the use of RTLS to improve the lives of residents and care providers in long-term care settings. More research related to RTLS use with cognitively impaired residents is required; this research should include longitudinal evaluation of end-to-end implementations that are developed using scientific theory and rigorous analysis of the functionality, efficiency, and effectiveness of these systems. Future research is required on the ethics of monitoring residents using RTLS and its impact on the privacy of residents and health care workers.

(J Med Internet Res 2021;23(1):e22831) doi:10.2196/22831

KEYWORDS

assistive technology; real-time location system; long-term care; implementation science; dementia; Alzheimer disease; ambulatory monitoring; wearable technology; qualitative research

Introduction

Background

In Canada, approximately 87% of people living in residential long-term care homes, or nursing homes, have dementia or another type of cognitive disability [1]. This percentage is expected to increase as the number of people older than 65 years is predicted to grow by 68% over the next 20 years [2] and the number of people living in North America with dementia is anticipated to increase by 63% [3]. The subsequent increase in the number of older adults living with cognitive disabilities is driving health care decision makers to consider the use of automation and novel technologies in an effort to decrease the cost of care while improving safety, quality of care, and efficiency in long-term care [4].

Real-time location systems (RTLS) consist of a software app and reference points that detect and synthesize positioning data from wireless tags or transmitters attached to objects or people. RTLS are being adopted in health care with the goal of improving health and safety. Uses of RTLS include monitoring handwashing [5], fall prevention [6], tracking of individuals [5,7] and assets [8], enhancing independence [9,10], and more recently for collecting health data [6,11-13]. RTLS use tags or badges attached to patients or assets, receiver devices, and software to continuously and inconspicuously monitor activity in real time over wireless networks.

The rising interest in the use of RTLS in long-term care homes is related not only to the ability to remotely track the movement of individuals and assets [5] but also to its potential to gather clinically significant objective streams of data to augment care providers' subjective observation [6] and to increase safety and productivity in the face of an aging workforce [14,15]. RTLS are increasingly perceived as having the potential to enhance independence and improve the physical safety of residents living in long-term care while also reducing the widespread use of restrictive measures such as antipsychotic drugs, physical restraints, and secure units [16-18]. These *first-generation smart systems* [19] that use wireless geolocation have been successfully implemented in other industries such as insurance [20] and telecommunications [21] but present novel challenges when deployed in health care, particularly when they are also used to generate personal data from vulnerable people [22]. In fact, although the intent of care providers and organizations is to use RTLS to improve the lives of people living with dementia or other types of cognitive disabilities, their use may provide the opposite result by threatening residents' privacy and restricting their activity within the range of the technology [23]. Industry has leveraged the ability to collect, store, combine, and analyze large quantities of data about customers, despite data sets being anonymized. The disclosure of sensitive personal information, such as location data, poses reputational risk and risks to employability, insurability, and even civil liability. Moreover, the probability of harm increases with time and the frequency of data collection [24]. In health care, critical approaches to the implementation of monitoring technologies suggest the need for institutional accountability for any increased harm from their implementation, particularly for vulnerable populations [25]. Furthermore, there are recognized global standards around privacy for persons with disabilities, such as the United Nations Convention on the Rights of Persons with Disabilities, Article 22, that protects against privacy intrusions *regarding the place of residence or living arrangements* [26], which should also guide the implementation of monitoring technologies. Pursuant to recent criticism that the ethical standard of autonomy based on independence and noninterference is less relevant in long-term care [27], the introduction of RTLS into institutional settings requires a cautious approach that identifies all stakeholders' perspectives (eg, residents, residents' families, care providers, organizations, and society in general) and accounts for the benefits and risks to each that may result. Implementation of these technologies should only proceed when all risks have been accounted for and mitigated or when they are outweighed by agreed-upon aggregated benefits.

Objectives

The question as to whether RTLS are valid, reliable, accurate, and adoptable in long-term care settings, particularly for the purpose of optimizing independence and safety for residents living with dementia or other cognitive disabilities is not well reported. In this paper, we explore the factors influencing the implementation, adoption, and use of such systems by conducting a systematic review to identify related literature from academic and peer-reviewed journals. The study of health care interventions, particularly focusing on the understanding and evaluation of their implementation, is important for researchers and health care managers. The interaction of agents such as organizations, care providers, and residents, their contexts, and their processes (both established and introduced by an intervention or novel technology) must be anticipated or, at the very least, accommodated in a successful implementation [28,29]. The authors are part of a large team of researchers exploring whether clinically relevant physical, cognitive, and mental well-being information about residents living with dementia can be derived from geolocation data. Before introducing RTLS software to various pilot locations that provide residential long-term care for adults with dementia or other cognitive disabilities, the authors seek to better understand the factors influencing the success of RTLS implementation, particularly in long-term care settings (eg, residential long-term care homes, long-stay psychiatric care facilities).

Methods

We conducted a systematic review of the academic, peer-reviewed literature using the Joanna Briggs Institute (JBI) approach [30] developed to support evidence-based practice in health care settings [31]. The JBI methodology allows for the appraisal and integration of quantitative and qualitative forms of knowledge related to the research question so that findings from both methodologies may complement one another. This type of review strengthens the findings and generates more robust conclusions, thus making them more applicable to policy and practice [32]. Given the complexity and novelty of implementing and evaluating RTLS in long-term care settings, this methodology is especially suitable. Inclusion and exclusion criteria were prespecified, and procedural decisions were documented.

Data Sources and Searching

In collaboration with a Medical Library Information Specialist, we searched the electronic databases Embase (1974: week 25, 2019), CINAHL (1981: August 3, 2019), MEDLINE (1946: August 3, 2019), and PsycINFO (1967: August 3, 2019). Search strategies included subject headings and keywords related to 3 concepts that were combined using *AND OR*. The search concepts were (1) illnesses, diagnoses, and disorders associated with cognitive disability; (2) long-term care settings (eg, long-term care homes); and (3) RTLS (eg, GPS, sensor-based systems). The Ovid MEDLINE search strategy is given in Multimedia Appendix 1. We restricted searches to adult populations (18 years and older), and primary research studies and reviews published in English. The hand-searching process included reviewing relevant journals and references of included studies and searching Google Scholar to identify unindexed references; selected papers were included in the screening process. An updated search was conducted on all databases, and citations of selected papers, to April 16, 2020, resulting in the addition of one paper.

Study Selection

We included primary research studies and reviews that (1) used and/or described issues relating to RTLS, (2) focused on adults 18 years or older with cognitive disability, and (3) focused on long-term care homes or other types of residential facility settings. In the screening process, we excluded studies that (1) did not satisfy the inclusion criteria; (2) involved, or referred to, children or youth (17 years or younger); (3) focused on individuals without cognitive disabilities; or (4) were conference abstracts, editorials, and commentaries.

Before commencing the screening process, a calibration exercise was conducted to ensure reliability in correctly selecting paper for inclusion. This process entailed 4 researchers (YK, AG, JM, and AB) independently screening a random sample of the references. This process was repeated until unanimity was achieved; any conflicts were resolved through discussion. Titles and abstracts of all references were then screened independently for inclusion by at least 2 reviewers (YK and AG or JM). Conflicts were resolved via discussion or by a third reviewer. All references that were found to be eligible or potentially eligible underwent full-text screening to confirm eligibility, papers were screened independently by 3 reviewers (YK, AG, and JM), and conflicts were resolved by discussion.

Data Extraction and Quality Assessment

Data and information from publications were systematically extracted using a prepiloted data extraction form created in Microsoft Excel. Extracted information included publication information and type, methodology, theoretical or conceptual frameworks used, target population characteristics, study participant characteristics, institution information, and nature of RTLS. One reviewer independently extracted this information (YK), and the extracted data were reviewed by at least 1 other reviewer (JM or AG). All included publications were critically appraised for quality using the corresponding JBI quality assessment tool for the publication type and/or study design [33]. One reviewer independently assessed quality, and assessments were confirmed by at least 1 other reviewer (JM or AG).

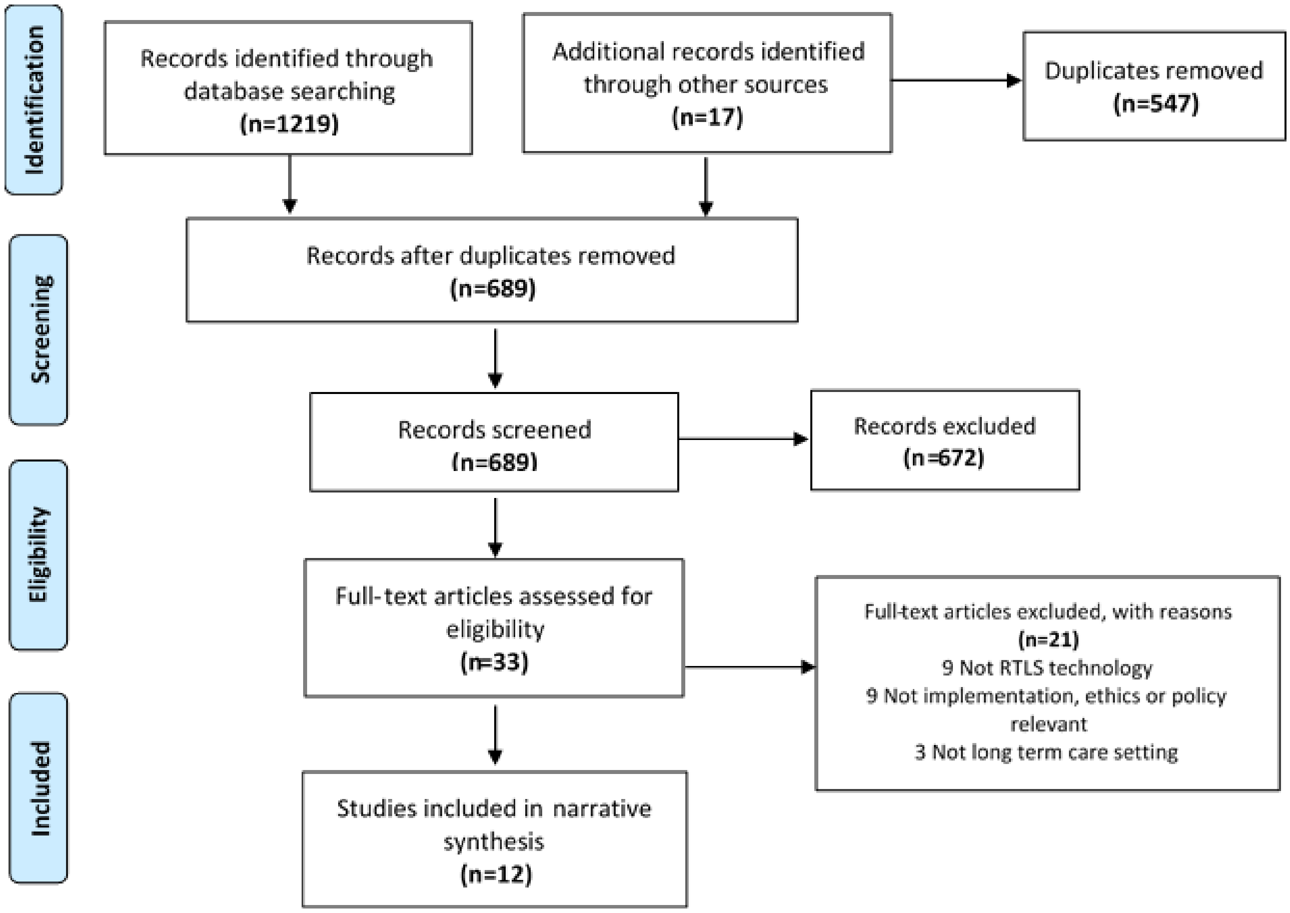
Data Analysis

Selected publications were imported into NVivo 12.0 software (QSR International) for data management and analysis. An initial set of categories were identified a priori during the scope determination phase of the review; these provided the basis for the initial round of descriptive coding (eg, resident, outcomes, risk). We progressively developed new concepts, and codes were then added, updated, or deleted and then later combined into larger themes. By applying constant comparison techniques, we iteratively revised the coding framework and simultaneously coded and analyzed the data [34,35]. This began with 3 authors (JM, AG, YK) independently reading each paper several times to identify links between them and conducting an initial round of coding using the initial set of codes and identifying new ones (eg, open coding). We then discussed and further refined the coding framework [36]. This was followed by several rounds of axial coding and selective coding comparing and contrasting findings across studies, identifying patterns, and conceptualizing broader themes that were then iteratively discussed among the authors (JM, AG, YK), with conflicts resolved via consensus [37].

Results

A total of 689 eligible references were identified through literature searches, and after screening, 12 publications were included in the review and are appended in Multimedia Appendix 2 [4,9,15,27,38-45]. Figure 1 shows a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram summarizing the results of the searching and screening process. All studies were published between 2010 and 2019. Most studies (7/12, 58%) were conducted in the Netherlands, and half of the total studies focused exclusively on persons living with dementia (6/12, 50%). A variety of study designs were used; 3 were quasi-experimental quantitative studies [38-40], 2 were concept mapping studies [41,42], 2 were qualitative case studies [4,43], 1 was a descriptive study [44], 1 was a qualitative descriptive study [9], 1 was an ethnographic study [45], and 2 were literature reviews (Niemeijer et al [45]; Oude Weernink et al [15]). Of the 12 publications, 2 were analyses of the same data set [4,43]. Of the publications examining the actual use, as opposed to the potential use, of RTLS (8/12, 67%), the most commonly used technologies were GPS and radio-frequency identification. Seven of the 12 publications collected qualitative data and/or feedback from individuals who would be tracked by the RTLS and other stakeholders interacting with the technology (eg, care providers); of these, 3 included data and feedback from residents. We conducted quality assessments of the included studies using the applicable JBI quality assessment tool. Quality assessments were performed independently by 1 review author (YK) and verified by a second review author (AG or JM). Most of the publications were considered good quality, except for the study by Bowen et al [44], which was rated as poor quality. None of the studies were excluded on the basis of quality assessments.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. RTLS: real-time locating system.



The selected studies were not overtly informed by implementation science theory [28]. Although normalization process theory [46] informed data collection for 1 study [43], its absence in the study design and analysis makes it difficult to extract meaning and better understand and control the barriers and facilitators that Greenhalgh et al [47] suggested actively affect the technology effectiveness. However, in our analysis of the collected data, we were able to identify 3 implementation-related themes (eg, barriers, enablers, agency and context) and several subthemes that may be useful for informing the execution of RTLS projects in residential care settings for adults living with cognitive disabilities.

Theme 1: Barriers to Implementation

Implementation barriers are those factors that, if present, will actively impede the adoption and use of an RTLS. They are aggregated under 3 categories: lack of motivation for engagement; technology ecosystem and infrastructure; and myths, stories, and shared understanding. All are relevant to the resident, frontline care provider, and/or organizational contexts.

Lack of Motivation for Engagement

As with other technologies, an RTLS is likely to be used less often and at a fraction of its capabilities; lack of care providers' and other stakeholders' engagement is considered a barrier to technology adoption and meaningful use [48,49]. Hall et al [4] noted that the most common justification for the use of monitoring technologies such as RTLS for persons living with cognitive disability is to enhance safety and quality of care while also achieving institutional aims of functional efficacy, increased security (or reduction of risks), and *reduction of staff burden*, and if these expectations are not met, it elicited responses of disappointment and disengagement. Unless there is a compelling rationale for the need to increase a resident's safety, frontline care providers appeared unwilling to support RTLS implementations [4]. Bowen et al [44] also found that care providers were less engaged when they viewed implementation as being conducted only for research purposes rather than for a permanent adoption that would affect their clinical routines. The persons who are the intended target for monitoring using RTLS in the selected papers are among the lowest adopters of novel technologies [50], consistent with Olphert and Damodaran's [51] *digitally disengaged* populations, who experience cognitive and physical challenges with the uptake and sustained use of digital technologies. Although this may be a generational phenomenon, all of the studies indicated that resident participants in the selected studies had limited technological familiarity or expectations with respect to the RTLS (eg, study by Hall et al [43]) and were not perceived as being motivated or capable of understanding the implications of these technologies beyond their having more freedom of movement as a result [4]. However, they were able to discern others' freedom as being different from their own as a result of others' use of RTLS or find new private spaces using it themselves [27].

The direct and indirect costs of RTLS represent a distinct barrier to organizational engagement and implementation in long-term care settings where incremental technology costs are scrutinized carefully for cost efficiencies such as reduced staffing costs [40] and complexity versus high value outcomes such as greater safety in comparison with the status quo [4,43]. Furthermore, contractual agreements between institutions and incumbent suppliers may limit access to novel technologies and prevent personalization of digital solutions that improve adoption [43]. With economies of scale favoring larger institutional competitors, there is a higher likelihood of historic or contractual procurement obligations, preventing the entrance of new competitors willing to adapt to organizational needs [4].

Care providers' and residents' lack of understanding of organizational strategic goals and value of the RTLS and the data it produces may also negatively affect their engagement [4,9,44]. This might explain care providers removing or not replacing removable tags on patients, not responding to alerts [4,44], and not replacing tag batteries [4]. Providers are also reluctant adopters of the technology when one of the direct or indirect goals is workforce surveillance, whether it potentially enhances organizational reputation or quality of care [43]. Managers acknowledged the autocratic implications of RTLS on both care providers and residents, as did the RTLS developers, although one study reported a lack of awareness among providers and family members of residents when discussing these ethical tensions [4,43].

All selected papers, except Bowen et al [44], who focused on what they described as *critical adoption issues that must be overcome for deployment* (p 191), identified ethical concerns related to the use of RTLS. Hall et al [43] pointed to the complexity and discordance of the ethical issues that arise when the locations in which RTLS are installed have multiple *identities*—as a workspace, a space for care delivery, and as a residential domicile. Niemeijer et al [45] focused on the ethical concerns regarding surveillance technologies internationally but presented these issues not as *barriers* to adoption of RTLS per se, but rather considerations for an *ethical debate* (p 1129). The *debate* occurs when persons living with cognitive disabilities are perceived as lacking decision-making capacity and when providers' and families' duties to safeguard residents' personal health and safety conflict with residents' right to privacy and dignity [52]. For this reason, we do not present ethical dilemmas that result from the use of RTLS as a barrier but will explore them in detail in the *Discussion* section.

Technology Ecosystem and Infrastructure Challenges

Underperforming technology is one of the greatest barriers to successful RTLS implementation [15,53]. The ecosystem into which it is introduced must support its functionality. For instance, RTLS that take a walled-garden approach to protect an intellectual property or distinctiveness compromise interoperability with other institutional applications or platforms and add unnecessary and often unrealistic complexity that negates the value of their use [43]. Other issues include lack of range and signal strength [40], tag battery life [4,15,40], loss of antenna strength and connection [40], and tags that are easily soiled or misplaced [44]. The functionality of an RTLS can present direct and unintended barriers to adoption. For instance, the ability to remove a system's wearable sensors is necessary for daily hygiene and maintenance purposes but presents an opportunity for these sensors to become lost—a catastrophic failure of this system, given its singular purpose to locate an object in time and space [44]. Some managers cited the technology's lack of robustness and inability to withstand the rigors of daily use as reasons to reject or reconsider its use [4]. Facility design is often not ideal for RTLS installations, where thick walls or crowded space for installation of cables or new equipment might thwart their optimal functioning [44]. The frequency of false-positive alarms has been identified as 1 of the top 10 hazards in medical device technology [54] and a workflow disruptor that can lead to health care provider error [55] and fatigue [56]. Numerous studies mentioned the burden of frequent alarms generated by RTLS as negative for care providers and residents but conversely noted that they raised providers' awareness of potentially risky incidents [4].

The lack of information in the papers related to location data storage, system security, and the specifics of data ownership and use was notable. The review paper by Oude Weernink et al [15] is the only one that identified system security (rather than security related to the safety of the individual wearer) as a potential barrier to the acceptance of RTLS. As the decision to use and place pervasive monitoring devices is not made by residents, the sacrifice of privacy for independence is not consciously made by residents. Niemeijer et al [45] suggest that discussions about data security and attitudes toward monitoring technologies are more skeptical in Europe than North America, perhaps reflecting the timing and adoption of the General Data Protection Regulation in 2016 regulating institutions' safe collection and transfer of individuals' private data. However, this was not evident in the selected literature.

Myths, Stories, and Shared Understanding

Insufficient training and clear communication preimplementation may contribute to the perception that use of RTLS supports the normative *blame culture* that arises when accidents occur in long-term care facilities [4,43]. Early care providers' and other stakeholders' engagement and instruction that discusses not only the functionality of the technology but also its intended benefits and anticipated challenges might address myths and disinformation [44] or potential resistance to RTLS use [15], all of which create barriers to implementation. Care providers were particularly worried about the potential for RTLS to be used for workplace oversight, although management in these studies appeared reluctant to acknowledge its active use in the supervision of care providers [4,43]. Inevitably, installation of an RTLS will affect existing routines and work practices [4], although they present less of a barrier to implementation when they fit with the technology's processes and functionality [44]. Although there are indications that RTLS may be useful in identifying traffic flow bottlenecks of providers or residents, helping to prevent medical errors, and reducing resident and operational risks [15], these systems are insufficient correctives to problematic workplace attitudes and cultural issues [43]. The myths that RTLS have no discernible negative effects manifested in a variety of areas such as reduced commitment to training and hasty implementations and are justified by its apparent simplicity [4].

Theme 2: Implementation Enablers

Implementation enablers are those factors that, if present, will actively facilitate the adoption and use of an RTLS. They are classified into 4 categories: understanding local workflows, policies, and technologies; usability and user-centered design; communication with care providers; and establishing policies, frameworks, governance, and evaluation.

Understanding Local Workflows, Policies, and Technologies

Bowen et al [44] described lessons learned where researchers and/or those implementing RTLS should ensure that they are aware of local and site-specific workflows, policies, technologies, and layout to ensure that the RTLS can be implemented effectively. Examples include ensuring that RTLS do not emit signals that might interfere with existing medical equipment and selecting RTLS wearables that are distinct from other types of wearables for ease of identification. In addition, trialing the technology before its implementation is key to identify and modify the technology to ensure that it does not potentially conflict or interfere with the existing workflows, policies, and other types of technology.

Usability and User-Centered Design

Many of the studies, both experimental and descriptive, suggest that the usability and acceptability of RTLS would be enhanced through the active engagement of all stakeholders, including residents, families, and care providers in implementation planning [15,40,42-45]. Bowen et al [44] found that engaging care providers and residents regarding the proposed RTLS and its components allowed them to modify the wearable component so that it was perceived as being less identifiable as a surveillance device by integrating a timepiece into the wristband. Hall et al [4,43] and Masciadri et al [40] suggested that designing an RTLS to be as unobtrusive as possible is an enabler, as it will be less noticeable to monitored residents. Niemeijer et al [41,45] added that implementation of RTLS should be person-specific and tailored to meet individual residents' needs and requirements and their families' preferences. Similarly, Masciadri et al [40] suggested that implementation should be context-specific and tailored to the preferences of stakeholders (eg, higher numbers of sensors for elevators and entrances for safety).

Communication With Care Providers

Communicating with, and providing training to, care providers regularly is suggested as enabling acceptance, compliance, understanding, and effective usage of RTLS [4,15,41,44]. However, communications with care providers seldom occurred through, or in concert with, leadership or management teams. Often, research teams would establish direct relationships and lines of communication with individual care providers, which improved their acceptance and understanding. Bowen et al [44] found that communicating and obtaining approval with/from leadership and management teams did not ensure that study information was communicated to all care providers in the home and suggested that researchers should make separate efforts to inform and accommodate care providers across different shift schedules [4,44]. Bowen et al [44] further suggested that conducting regular visits to implementation sites can help to address emerging concerns with providers and identify the root causes of challenges. Communication with care providers, as well as their training, should also be tailored to ensure that the aims, benefits, and usage of the technology, how it fits in with the values and practices of the home, and its functional usage [4,15,41] are clear.

Establishing Policies, Frameworks, Governance, and Evaluation

Ensuring care providers and organizations are aware of whose responsibility it is to maintain the technology and to whom they report any issues would increase their confidence [41,42]. Similarly, relevant laws and policies with respect to privacy protection should be clearly defined and articulated to all stakeholders [42]. Contingency and emergency planning should also be discussed so that all are aware of how any potential failings will be addressed and by whom, as technologies become embedded in usual care [42]. Regular and clear evaluation of the monitoring technology should occur, including evaluating the quality of life of residents who are being monitored, to help guide decision making regarding whether the technology should be used. This decision-making process should be clearly described and communicated to everyone involved [41,43,45].

Theme 3: Agency and Context

The implementation of an RTLS in any setting is influenced by what people do (agentic contributions) and the relationship between agents (individuals and groups) and their contexts (spatial, organizational, and normative). In long-term care homes, this is made more complex by the tensions and opportunities for actions that arise as a result of the goals and values of different stakeholders or groups of agents (eg, residents, family members, care providers, and organizations), their relationship with each other, and their relative decision-making power with respect to the adoption and use of the RTLS. Despite this complexity, the collected papers rarely acknowledge these tensions, and for the most part, explore the motivations and actions of care providers and organizations in isolation, largely divorced from a consideration of the relationship between them, with others, or with the context. This is problematic, as research in other congregate settings for older adults on monitoring technologies demonstrates that the failure to account for the mismatch in priorities between stakeholders may lead to resistance and discontinuation [57].

Residents

Although residents living with dementia are identified as the primary agent on whose behalf RTLS are implemented in long-term care settings, they are not the people who are driving their implementation. Little research has been done on the perceptions and experiences of residents with this type of technology, and they are generally perceived as being passive recipients of care. Furthermore, although the main justification for the need for an RTLS is that it will improve the independence and safety of residents living with dementia (eg, increased freedom of movement, prevention of harm), almost no research has explored this in practice. The available research focuses on the use of restraints and demonstrates that there is no difference in positive affect between residents with high activity of daily living (ADL) dependency who are physically restrained and residents monitored with an RTLS [39] and suggests that the introduction of RTLS does not lead to a reduction in the use of physical restraints [9].

Residents are described as being generally accepting of the use of RTLS [44], perhaps because they do not notice the technology and/or understand exactly how it works or influences their movements or interactions with providers [4,44]. Their occasional resistance to these technologies is dismissed as being because of a lack of understanding, which is attributed to their cognitive disability or inadequate technological design, rather than being accepted as a valid expression of choice. For example, Bowen et al [44] reported that residents who objected to these technologies did so because they mistakenly believed that the sensors could videotape, record their conversations, or restrict their movements in real time. Similarly, Hall et al [4] and Bowen et al [44] reported that explaining the purpose of the technology to residents—monitoring their movements remotely to enhance freedom of movement—could lead to rejection of an RTLS if the residents were not already aware that their movements were restricted or monitored otherwise. Residents’ removal of tags or wearable components of RTLS was attributed to poor design esthetics and/or uncomfortable placement (eg, male residents who refused to wear a pink wearable or perceived being tagged with a bracelet as stigmatizing) rather their objection to their purpose [4,44].

Family Members

Available research suggests that family members are treated by direct care staff and organizations as key decision makers with respect to implementation of RTLS, regardless of whether they act as a substitute decision maker or have formal power of attorney [58]. Family members generally endorse the use of monitoring technologies for residents if they believe that it enhances residents’ physical safety by either supporting earlier detection of risk of injury (eg, fall prevention, altercation with another resident) or risk of neglect or abuse from staff [40,41,44,45]. They also prioritize the physical safety of residents over the potential risk that these technologies may pose to their privacy and/or autonomy, with some perceiving technological monitoring without the use of video cameras as being more privacy-protecting than continuous in-person observation [41,45].

Care Providers

Similar to family members, providers perceived RTLS as being valuable components of care that support their responsibility to safeguard residents, and most research focused on their use of these systems to enhance in-person monitoring [4,9,43]. A key cited benefit of RTLS, generally supported by providers, was the ability to locate residents or monitor their movement remotely in real time [15,40,44]. Care providers, however, were less enthusiastic about the use of RTLS by the organization to locate or to monitor their activities and expressed worry that these technologies might be used to sanction them. For the most part, researchers did not explore this topic in depth, with some suggesting that providers’ worries were unfounded and because of their lack of understanding of the purpose of implementing an RTLS (eg, Bowen et al [44] called it a *myth*). Lack of trust in the technology because of its nascent development and ethical implications of using it with vulnerable populations can result in disengagement by care providers and residents either through nonuse or by use that ignores the risks and may result in overconfidence in its capabilities [45].

Organization

The main stakeholders behind implementations of RTLS appear to be senior leadership or management [4,43], who perceive these technologies as enhancing organizational protection from risk and liability (eg, prevention of injury to residents, defense against allegations of negligence) or as leading to cost savings (eg, reduction in providers, monitoring provider performance) [4,40,42-45]. These stakeholders also suggested that continuous data collection using RTLS may be useful for mitigating family members’ potential concerns about residents or as protection against complaints and litigation regarding neglect or abuse [40,44]. However, no studies have explored whether these benefits are realized in practice or whether organizations use RTLS in this manner. Management does, however, use RTLS to monitor providers’ performance in both covert and overt ways and perceives this as permissible based on organizations’ duty of care to residents [4,43]. In general, management considered the risks of RTLS to be less of a barrier to adoption than the risks to the institutional reputation of not meeting their duty of care through adequate facilities, services, and *innovation through people* [43].

Discussion

With the anticipated aging of the population in many western countries [59-61], there is increasing interest in the potential of using technologies to improve the health, safety, and quality of life in long-term care homes while reducing the cost of care delivery. Our review of the scientific literature related to the use of RTLS to monitor individuals living with dementia or other types of cognitive disability in residential care settings reveals scant evidence of implementation science theory to inform and optimize outcomes, with the exception of 1 study [4,43]. Implementation science advocates for a clear understanding and resolution of issues related to novel technologies (in their broadest sense, including artifacts such as hardware, software, processes, and policies) and the context and setting in which it is being applied [62,63].

Much of the justification for the use of RTLS or other types of surveillance technologies in long-term care focuses on its potential to enhance the quality of life and physical safety of residents with cognitive disabilities, in particular through the avoidance of segregation or physical restraints [42]. Te Boekhorst et al [39] and Zwijsen et al [9] found that care providers perceive the use of surveillance technology as an *intermediate measure* for use before physical restraints, and participants in a study by Niemeijer et al [42] reported that although surveillance technology could prevent some forms of *freedom restriction*, it too can impede movement. The limited available evidence does not demonstrate a reduction in the use of restraints in practice, although it is unclear whether this was because of the misapplication of these technologies. For example, although Te Boekhorst et al [39] did not find that RTLS enhanced the quality of life for residents, they suggested that the restricted movement of residents with high ADL dependency might negate any independence benefits ascribed to the use of RTLS. Similarly, Zwijsen et al [9] suggested that an RTLS was not seen as a replacement for restraints because although it supported ubiquitous monitoring, it did not prevent risk or replace the need for providers to respond to risky situations [4]. Despite the suggestion that RTLS might reduce or replace in-person care (which is often an expressed concern of providers), the proper use and monitoring of the technology can be equally time-consuming [45]. The flood of data that require monitoring and action was seen as a potential deterrent to its adoption, in particular because of the lack of expertise in transforming data into clinically useful information in practice [4]. Although there is some value in improving communications about the technology and processes, Bowen et al [43] demonstrated that this does not translate into improved trust in the technology or management or adherence to protocols. Trust was also challenged when the RTLS was unreliable, requiring workarounds and additional provider time [4].

Technologies are increasingly *integrated* with humans, from wearables such as RTLS to biomedical implants such as pacemakers. As Latour and Venn [64] have pointed out, this intermingling means that technologies influence the behaviors and processes of humans and their institutions and vice versa. The features of some technologies, such as ubiquitous surveillance, may, as a result, be perceived as ethically untenable and unalterable and, thus, may not be appropriate for use in some populations or settings. Although the focus of this study is not on the ethics of RTLS implementation per se, half of the selected papers in this study report ethical issues to a greater or lesser extent. The presence of ethical arguments acknowledges that the use of RTLS introduces tensions between the values, goals, and autonomy of different stakeholders (eg, providers or organizations and residents) with respect to the use of RTLS, particularly when they are used in the care of persons living with dementia or other types of cognitive disabilities in long-term care settings [9,45,65]. This population is considered vulnerable, and both providers and organizations have a duty of care to them, suggesting an asymmetry of dependency and dignity that must be acknowledged and resolved before implementation of surveillance technology. Niemeijer et al [41] suggest that there is an *inherent duality… rooted in the moral conflict between safety and freedom* (p 303), where autonomy is offered in return for surveillance using RTLS [15]. They further point out that older adults in particular may feel that they have to accept an RTLS and sacrifice personal privacy for independence in a type of digital quid pro quo. Furthermore, family members and substitute decision makers favor the use of RTLS to enhance the physical safety of residents and perceive the potential risks to privacy as less important [4,42,43]. Although there is limited research on the experiences of persons living with dementia with RTLS, it is concerning that the available research suggests that their refusal of RTLS may not be respected [4,44]; resolving this dilemma is an ethical imperative to avoid coercive practices and respect resident autonomy. This is supported by recent efforts to develop and implement new technologies by integrating the ethical values and priorities of stakeholders into design and development [66,67].

Recommendations

Health care organizations are bound by workplace safety and data privacy laws and regulations that struggle to keep up with novel technologies [68]. The literature suggests that care providers and residents must rely on the culture and values of each individual organization to guide the adoption of RTLS. Our analysis suggests that there is interest among providers and organizations to better integrate nascent RTLS into work routines that improve safety and quality of care. However, there are numerous barriers to their effective use, including lack of engagement, trust, shared understanding of goals, and a reliable and appropriate RTLS that meets the needs of residents and care settings. A variety of factors to improve implementation of RTLS in the long-term care setting should be considered, including integration with local workflows, policies, and technologies and ensuring that care providers, residents, and families are involved, and if possible, leading that process [65,69]. Given the financial constraints in this sector, policy makers should consider the creation of a variety of incentives to encourage the use of technology implementation best practices, such as improved staff training and technology infrastructure development, to optimize outcomes and impact on care quality [70].

Decision makers must also be aware that daily interactions with care providers are often the primary source of social contact for many residents, and thus, the replacement of human contact with surveillance technology may unintentionally increase social isolation [45]. Where possible, institutions should select a flexible, interoperable RTLS that allows customization to suit individual residents, settings, and technological ecosystems [43], understanding that resident and operational needs change over time [27]. Technology developers need to be sensitive to the financial constraints that dominate this sector [43] and aim to build technologies that are elegant in their simplicity and easy to use and focus on solving issues at a lower total cost of ownership (eg, financial, infrastructure, effort, privacy) [40] to institutions, providers, and residents. Issues such as poor battery life, lost tags, and interoperability are solvable with current technology [15]. Researchers are well advised to devise RTLS research protocols and pilot studies that prompt all relevant stakeholders to engage meaningfully with the technology.

The ethical use of RTLS in long-term care settings is similar to other decisions made by providers and management, in that it requires an assessment of the potential benefits and harms, both direct and indirect, to all stakeholders involved. However, this is complicated, given that persons living with dementia or other cognitive disabilities are not assessing the possible risks and benefits independently. Many preferred to defer to family members and other proxies rather than support residents in making independent choices about the use of monitoring technology [4]. Although substitute decision makers are legally mandated to make choices on behalf of residents who lack capacity, this does not absolve them, researchers, and care providers from asking whether the use of an RTLS is ethical, whether it is something residents would want, and what constitutes an acceptable risk as a result of their implementation. It is clear that the use of monitoring technologies with individuals who are nonverbal or have cognitive disabilities, to address issues such as mobility and independence and improve predictive clinical diagnostic capabilities, has preceded our understanding of the rights, risks, and unintended consequences of their use in long-term care. A complete assessment of these issues is recommended before any implementation, drawing on a growing body of work exploring ethical design and implementation of technologies for vulnerable populations [71,72]. Furthermore, it is optimal to anticipate conflicts of interest, such as incidental monitoring of care providers, before they become barriers to a successful implementation [45], and to explore other mechanisms for workforce management [43]. Although this study focused primarily on monitoring residents with RTLS, care providers were also monitored either directly with tags or indirectly through the imputation of activity related to residents, such as the time it took them to respond to an alarm. Although RTLS may offer data to help assess and optimize clinical workflows, automatically open and lock doors, and improve operations [15], it also has the potential to reduce employee privacy and may negatively affect care provider recruitment and retention [43] in a sector that already struggles to find qualified staff [73].

Limitations

Although the quality of selected papers was acceptable, the limited number of studies and their small sample sizes reflective of qualitative research methodologies and exploratory research suggest that generalization of our results and recommendations for future research should be limited to similar population samples and sectors. Furthermore, 3 of the studies that were included in this review [38-40] contributed a limited amount of data and information to the main research objective of examining factors influencing the implementation, adoption, and use of RTLS.

Conclusions and Future Research

There is a striking lack of evidence to support the justification and implementation of RTLS to improve the quality of life of residents and work of care providers in long-term care settings. More research related to RTLS use with individuals with cognitive disabilities is required and should include longitudinal evaluations of end-to-end implementations that are theoretically informed and include rigorous analysis of functionality, efficiency, and effectiveness in improving outcomes that are important to all stakeholders involved [4,42]. Empirical studies that rigorously evaluate the practical utility and adoption of RTLS and their related processes into a controlled environment, the value of customization to the requirements of individual residents, and technology infrastructure versus one-size-fits-all adoptions are also required to advance our understanding of their utility.

The use of RTLS to support workflow efficiencies, manage person-to-person contact, and collect clinical data for use in diagnosis and therapeutics is largely unexplored and offers opportunities for future research and use. The workload and care provider capacity for real-time monitoring and data management and analysis for optimal use and outcomes must be made explicit and included in cost-benefit analyses that precede the purchase and adoption of these systems. Furthermore, training for the operation and use of RTLS will require incremental skills training and increased staffing levels, at least in the short term, in a sector where availability of resources and high workloads are already problematic issues. Finally, ethical considerations related to monitoring residents with RTLS, and also directly or indirectly their care providers, are acknowledged but not settled and require further empirical research.

Acknowledgments

This study was funded by AGE-WELL (Aging Gracefully across Environments using technology to support Wellness, Engagement and Long Life) Network, which was funded by the Government of Canada’s Networks of Centres of Excellence Program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ovid MEDLINE search strategy.
[DOCX File , 18 KB - jmir_v23i1e22831_app1.docx]

Multimedia Appendix 2
A summary of selected studies.
[DOCX File , 27 KB - jmir_v23i1e22831_app2.docx]

References

1. Policy changes and educational supports help spur a decrease in inappropriate use of antipsychotics and restraints Internet. Canadian Institute for Health Information. URL: https://www.cihi.ca/en/dementia-in-canada/dementia-across-the-health-system/dementia-in-long-term-care#:~:text=Within [accessed 2020-12-23]
2. Dementia in Canada: Summary Internet. Canadian Institute for Health Information. URL: https://www.cihi.ca/en/dementia-in-canada/dementia-in-canada-summary [accessed 2020-12-23]
3. Alzheimer's Disease International, Bupa. Current and future cost and prevalence of Alzheimer's disease and other dementias. Alzheimer's Disease International (ADI) and Bupa. Illinois, USA,; 2013. URL: https://www.alz.co.uk/sites/default/files/pdfs/dementia-in-the-americas-ENGLISH.pdf [accessed 2021-01-01]
4. Hall A, Wilson C, Stanmore E, Todd CA. Implementing monitoring technologies in care homes for people with dementia: a qualitative exploration using Normalization Process Theory. Int J Nurs Stud 2017 Jul;72:60-70 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.04.008] [Medline: 28494333]
5. van Hoof J, Verboor J, Oude Weernink C, Sponselee A, Sturm J, Kazak J, et al. Real-time location systems for asset management in nursing homes: an explorative study of ethical aspects. Information 2018 Apr 07;9(4):80. [doi: 10.3390/info9040080]
6. Bowen ME, Crenshaw J, Stanhope SJ. Balance ability and cognitive impairment influence sustained walking in an assisted living facility. Arch Gerontol Geriatr 2018 Jul;77:133-141. [doi: 10.1016/j.archger.2018.05.004]
7. Bowen ME, Kearns W, Crenshaw JR, Stanhope SJ. Using a real-time locating system to measure walking activity associated with wandering behaviors among institutionalized older adults. J Vis Exp 2019 Feb 08(144):-. [doi: 10.3791/58834] [Medline: 30799867]
8. Bykova A. Institutes of innovative development: their role in regional clusters. Ekon Ann 2011;56(190):59-76. [doi: 10.2298/eka1190059b]
9. Zwijsen SA, Depla MF, Niemeijer AR, Francke AL, Hertogh CM. Surveillance technology: an alternative to physical restraints? A qualitative study among professionals working in nursing homes for people with dementia. Int J Nurs Stud 2012 Feb;49(2):212-219. [doi: 10.1016/j.ijnurstu.2011.09.002]
10. Robinson L, Hutchings D, Corner L, Finch T, Hughes J, Brittain K, et al. Balancing rights and risks: conflicting perspectives in the management of wandering in dementia. Health Risk Soc 2007 Oct 16;9(4):389-406. [doi: 10.1080/13698570701612774]
11. Akl A, Taati B, Mihailidis A. Autonomous unobtrusive detection of mild cognitive impairment in older adults. IEEE Trans Biomed Eng 2015 May;62(5):1383-1394 [FREE Full text] [doi: 10.1109/TBME.2015.2389149] [Medline: 25585407]
12. Bowen ME, Rowe M. Intraindividual changes in ambulation associated with falls in a population of vulnerable older adults in long-term care. Arch Phys Med Rehabil 2016 Nov;97(11):1963-1968. [doi: 10.1016/j.apmr.2016.05.013] [Medline: 27296900]
13. Jansen C, Diegelmann M, Schnabel E, Wahl H, Hauer K. Life-space and movement behavior in nursing home residents: results of a new sensor-based assessment and associated factors. BMC Geriatr 2017 Jan 28;17(1):36 [FREE Full text] [doi: 10.1186/s12877-017-0430-7] [Medline: 28129741]
14. Boulos M, Berry G. Real-time locating systems (RTLS) in healthcare: A condensed primer. International Journal of Health Geographics. 2012. URL: https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-11-25 [accessed 2020-12-23]
15. Oude Weernink C, Felix E, Verkuijlen P, Dierick-van Daele A, Kazak J, van Hoof J. Real-time location systems in nursing homes: state of the art and future applications. J Enabling Technol 2018 Jun 18;12(2):45-56. [doi: 10.1108/jet-11-2017-0046]
16. Ryden MB, Feldt KS, Oh HL, Brand K, Warne M, Weber E, et al. Relationships between aggressive behavior in cognitively impaired nursing home residents and use of restraints, psychoactive drugs, and secured units. Arch Psychiatr Nurs 1999 Aug;13(4):170-178. [doi: 10.1016/s0883-9417(99)80003-x] [Medline: 10478495]
17. Dementia in long-term care. Canadian Institute for Health Information. 2018. URL: https://www.cihi.ca/en/dementia-in-canada/dementia-care-across-the-health-system/dementia-in-long-term-care [accessed 2021-01-01]
18. Lachance C, Wright MD. Avoidance of Physical Restraint Use among Hospitalized Older Adults: A Review of Clinical Effectiveness and Guidelines. Ottawa, Canada: CADTH; Feb 28, 2019:1-22.
19. Sassone A, Grosso M, Poncino M, Macii E. Smart electronic systems: an overview. In: Smart Systems Integration and Simulation. Switzerland: Springer, Cham; 2016:5-21.
20. Insurance and Technology: Evolution and Revolution in a Digital World Internet. Morgan Stanley Research. 2014. URL: http://www.the-digital-insurer.com/wp-content/uploads/2014/10/372-evolution-revolution-how-insurers-stay-relevant-digital-world.pdf [accessed 2020-12-23]
21. Rost M, Cramer H, Belloni N, Holmquist L. Geolocation in the mobile web browser. In: UbiComp '10 Adjunct. 2010 Presented at: Ubicomp '10: The 2010 ACM Conference on Ubiquitous Computing; September, 2010; Copenhagen Denmark. [doi: 10.1145/1864431.1864468]

22. Wood W, Bennett AV, Basch E. Emerging use of patient generated health data in clinical research. Mol Oncol 2014;9:10. [doi: 10.1016/j.molonc.2014.08.006]
23. Bennett B, McDonald F, Beattie E, Carney T, Freckelton I, White B, et al. Assistive technologies for people with dementia: ethical considerations. Bulletin of the World Health Organization. Brisbane: World Health Organization; 2017 May 12. URL: https://www.who.int/bulletin/volumes/95/11/16-187484.pdf [accessed 2021-01-01]
24. Altman M, Wood A, O'Brien D, Gasser U. Practical approaches to big data privacy over time. Int Data Priv Law 2018;8(1):51. [doi: 10.1093/idpl/ipx027]
25. Grigorovich A, Kontos P. Towards responsible implementation of monitoring technologies in institutional care. Gerontologist 2020 Sep 15;60(7):1194-1201 [FREE Full text] [doi: 10.1093/geront/gnz190] [Medline: 31958118]
26. Convention on the Rights of Persons with Disabilities (CRPD). United Nations. 2016. URL: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html#Fulltext [accessed 2020-12-24]
27. Niemeijer AR, Depla MF, Frederiks BJ, Hertogh CM. The experiences of people with dementia and intellectual disabilities with surveillance technologies in residential care. Nurs Ethics 2015 May;22(3):307-320. [doi: 10.1177/0969733014533237] [Medline: 24913545]
28. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016 Dec 19;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
29. May C. Towards a general theory of implementation. Implement Sci 2013 Feb 13;8(1):18 [FREE Full text] [doi: 10.1186/1748-5908-8-18] [Medline: 23406398]
30. Introduction to JBI Systematic reviews. Joanna Briggs Institute. URL: https://wiki.joannabriggs.org/display/MANUAL/1.1+Introduction+to+JBI+Systematic+reviews [accessed 2020-12-23]
31. Jordan Z, Lockwood C, Munn Z, Aromataris E. The updated Joanna Briggs Institute Model of Evidence-Based Healthcare. Int J Evid Based Healthc 2019;17(1):58-71. [doi: 10.1097/xeb.0000000000000155]
32. Östlund U, Kidd L, Wengström Y, Rowa-Dewar N. Combining qualitative and quantitative research within mixed method research designs: a methodological review. Int J Nurs Stud 2011 Mar;48(3):369-383 [FREE Full text] [doi: 10.1016/j.ijnurstu.2010.10.005] [Medline: 21084086]
33. Joanna Briggs Institute. JBI Critical Appraisal Checklist for Qualitative Research. JBI Manual for Evidence Synthesis. URL: https://wiki.joannabriggs.org/display/MANUAL/Appendix+2.1%3A+JBI+Critical+Appraisal+Checklist+for+Qualitative+Research [accessed 2020-12-23]
34. Maykut P, Morehouse R. Beginning Qualitative Research: A philosophic Practice Guide. London, UK: The Falmer Press; 1994:A.
35. Parry K. Constant Comparison. In: The SAGE Encyclopedia of Social Science Research Methods. Oaks, CA: Sage Publications Inc; 2004:181.
36. Lockyer S. Coding qualitative data. In: The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: Sage Publications Inc; 2004:138.
37. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. California: Sage Publications Inc; 1998.
38. Yayama S, Yamakawa M, Suto S, Greiner C, Shigenobu K, Makimoto K. Discrepancy between subjective and objective assessments of wandering behaviours in dementia as measured by the Algase Wandering Scale and the Integrated Circuit tag monitoring system. Psychogeriatrics 2013 Jun 12;13(2):80-87. [doi: 10.1111/psyg.12011]
39. te Boekhorst S, Depla MF, Francke AL, Twisk JW, Zwijsen SA, Hertogh CM. Quality of life of nursing-home residents with dementia subject to surveillance technology versus physical restraints: an explorative study. Int J Geriatr Psychiatry 2012 May 30;28(4):356-363. [doi: 10.1002/gps.3831]
40. Masciadri A, Comai S, Salice F. Wellness assessment of Alzheimer's patients in an instrumented health-care facility. Sensors 2019 Aug 22;19(17):3658. [doi: 10.3390/s19173658] [Medline: 31443505]
41. Niemeijer AR, Frederiks BJ, Depla MF, Legemaate J, Eefsting JA, Hertogh CM. The ideal application of surveillance technology in residential care for people with dementia. J Med Ethics 2011 Feb 02;37(5):303-310. [doi: 10.1136/jme.2010.040774]
42. Niemeijer A, Frederiks B, Depla M, Eefsting J, Hertogh C. The place of surveillance technology in residential care for people with intellectual disabilities: is there an ideal model of application. J Intellect Disabil Res 2013 Mar;57(3):201-215. [doi: 10.1111/j.1365-2788.2011.01526.x] [Medline: 23379860]
43. Hall A, Brown Wilson C, Stanmore E, Todd C. Moving beyond 'safety' versus 'autonomy': a qualitative exploration of the ethics of using monitoring technologies in long-term dementia care. BMC Geriatr 2019 May 24;19(1):145 [FREE Full text] [doi: 10.1186/s12877-019-1155-6] [Medline: 31126240]
44. Bowen ME, Wingrave CA, Klanchar A, Craighead J. Tracking technology: lessons learned in two health care sites. Technol Health Care 2013;21(3):191-197. [doi: 10.3233/THC-130738] [Medline: 23792793]
45. Niemeijer AR, Frederiks BJ, Riphagen II, Legemaate J, Eefsting JA, Hertogh CM. Ethical and practical concerns of surveillance technologies in residential care for people with dementia or intellectual disabilities: an overview of the literature. Int Psychogeriatr 2010 Nov;22(7):1129-1142. [doi: 10.1017/S1041610210000037] [Medline: 20199699]

46. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
47. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
48. Weeks DL, Keeney BJ, Evans PC, Moore QD, Conrad DA. Provider perceptions of the electronic health record incentive programs: a survey of eligible professionals who have and have not attested to meaningful use. J Gen Intern Med 2015 Jan;30(1):123-130 [FREE Full text] [doi: 10.1007/s11606-014-3008-5] [Medline: 25164087]
49. Heisey-Grove D, Danehy L, Consolazio M, Lynch K, Mostashari F. A national study of challenges to electronic health record adoption and meaningful use. Med Care 2014 Feb;52(2):144-148. [doi: 10.1097/MLR.0000000000000038] [Medline: 24309669]
50. UK: smartphone ownership by age from 2012-2017 Internet. Statista. 2017. URL: https://www.statista.com/statistics/271851/smartphone-owners-in-the- [accessed 2020-12-23]
51. Olphert W, Damodaran L. Older people and digital disengagement: a fourth digital divide? Gerontology 2013;59(6):564-570. [doi: 10.1159/000353630] [Medline: 23969758]
52. Plastow NA. Is Big Brother watching you? Responding to tagging and tracking in dementia care. Br J Occup Ther 2006 Nov 01;69(11):525-527. [doi: 10.1177/030802260606901106]
53. Fisher JA, Monahan T. Evaluation of real-time location systems in their hospital contexts. Int J Med Inform 2012 Oct;81(10):705-712. [doi: 10.1016/j.ijmedinf.2012.07.001] [Medline: 22857790]
54. ECRI. ECRI Top 10 Health Technology Hazards for 2020. J Radiol Nurs 2020 Mar;39(1):6. [doi: 10.1016/s1546-0843(20)30009-2]
55. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
56. Schikhof Y, Mulder I, Choenni S. Who will watch (over) me? Humane monitoring in dementia care. Int J Hum Comput Stud 2010 Jun;68(6):410-422. [doi: 10.1016/j.ijhcs.2010.02.002]
57. Berridge C. Active subjects of passive monitoring: responses to a passive monitoring system in low-income independent living. Ageing Soc 2017 Mar;37(3):537-560 [FREE Full text] [doi: 10.1017/S0144686X15001269] [Medline: 28239211]
58. A Guide to the Substitute Decisions Act Internet. Ministry of the Attorney General. 2000. URL: https://www.attorneygeneral.jus.gov.on.ca/english/family/pgt/pgtsda.pdf [accessed 2020-12-23]
59. Peek ST, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016 Mar;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
60. Bodenheimer T. High and rising health care costs. Part 1: seeking an explanation. Ann Intern Med 2005 May 17;142(10):847-854 [FREE Full text] [doi: 10.7326/0003-4819-142-10-200505170-00010] [Medline: 15897535]
61. Carstairs S, Keon W. Canada’s Aging Population: Seizing the Opportunity. Special Senate Committee on Aging Final Report. Ottawa: Senate, Canada; 2009. URL: https://sencanada.ca/content/sen/Committee/402/agei/rep/AgingFinalReport-e.pdf [accessed 2021-01-01]
62. Schoville RR, Titler MG. Guiding healthcare technology implementation: a new integrated technology implementation model. Comput Inform Nurs 2015 Mar;33(3):99-107. [doi: 10.1097/CIN.0000000000000130] [Medline: 25799235]
63. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
64. Latour B, Venn C. Morality and Technology. Sage Publications. 2017 Jan 13. URL: https://journals.sagepub.com/doi/pdf/10.1177/026327602761899246 [accessed 2021-01-01]
65. Bourbonnais A, Rousseau J, Lalonde M, Meunier J, Lapierre N, Gagnon M. Conditions and ethical challenges that could influence the implementation of technologies in nursing homes: A qualitative study. Int J Older People Nurs 2019 Dec 05;14(4):e12266 [FREE Full text] [doi: 10.1111/opn.12266] [Medline: 31475466]
66. Cummings ML. Integrating ethics in design through the value-sensitive design approach. Sci Eng Ethics 2006 Oct;12(4):701-715. [doi: 10.1007/s11948-006-0065-0] [Medline: 17199145]
67. Pacifico Silva H, Lehoux P, Miller FA, Denis J. Introducing responsible innovation in health: a policy-oriented framework. Health Res Policy Syst 2018 Sep 10;16(1):90. [doi: 10.1186/s12961-018-0362-5] [Medline: 30200985]
68. Maguire C. The struggle to keep up with data privacy regulations. Law Journal Newsletters. URL: https://www.lawjournalnewsletters.com/2018/04/01/the-struggle-to-keep-up-with-data-privacy-regulations/?slreturn=20200730165959 [accessed 2020-12-23]
69. Bourbonnais A, Ducharme F, Landreville P, Michaud C, Gauthier M, Lavallée MH. An action research to optimize the well-being of older people in nursing homes: challenges and strategies for implementing a complex intervention. J Appl Gerontol 2020 Feb;39(2):119-128. [doi: 10.1177/0733464818762068] [Medline: 29504489]

70. Ko M, Wagner L, Spetz J. Nursing home implementation of health information technology: review of the literature finds inadequate investment in preparation, infrastructure, and training. Inquiry 2018;55:46958018778902 [FREE Full text] [doi: 10.1177/0046958018778902] [Medline: 29888677]
71. Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical design of intelligent assistive technologies for dementia: a descriptive review. Sci Eng Ethics 2018 Aug;24(4):1035-1055. [doi: 10.1007/s11948-017-9976-1] [Medline: 28940133]
72. Fields LM, Calvert JD. Informed consent procedures with cognitively impaired patients: a review of ethics and best practices. Psychiatry Clin Neurosci 2015 Aug;69(8):462-471 [FREE Full text] [doi: 10.1111/pcn.12289] [Medline: 25756740]
73. Stone L, Keller J. Ontario relying on volunteers to help staffing shortage in long-term care as feds release new guidelines. Globe Mail Internet Toronto, Canada 2020:- [FREE Full text] [doi: 10.14361/9783839438053-023]

Abbreviations

ADL: activities of daily living
JBI: Joanna Briggs Institute
RTLS: real-time locating system

Edited by G Eysenbach; submitted 24.07.20; peer-reviewed by J He, IV George; comments to author 28.08.20; revised version received 31.08.20; accepted 29.10.20; published 20.01.21.

Please cite as:
Grigorovich A, Kulandaivelu Y, Newman K, Bianchi A, Khan SS, Iaboni A, McMurray J
Factors Affecting the Implementation, Use, and Adoption of Real-Time Location System Technology for Persons Living With Cognitive Disabilities in Long-term Care Homes: Systematic Review
J Med Internet Res 2021;23(1):e22831
URL: http://www.jmir.org/2021/1/e22831/
doi:10.2196/22831
PMID:33470949

Review

Determining the Intellectual Structure and Academic Trends of Smart Home Health Care Research: Coword and Topic Analyses

Hyo-Jin Kang[1], BSc, MBA, PhD; Jieun Han[2], BSc, PhD; Gyu Hyun Kwon[2], BSc, MSc, PhD

[1]Department of Service Design Engineering, Sungshin Women's University, Seoul, Republic of Korea

[2]Graduate School of Technology and Innovation Management, Hanyang University, Seoul, Republic of Korea

Corresponding Author:
Gyu Hyun Kwon, BSc, MSc, PhD
Graduate School of Technology and Innovation Management
Hanyang University
222 Wangsimni-ro
Seongdong-gu
Seoul, 04763
Republic of Korea
Phone: 82 2 2220 2414
Email: ghkwon@hanyang.ac.kr

Abstract

Background: With the rapid development of information and communication technologies, smart homes are being investigated as effective solutions for home health care. The increasing academic attention on smart home health care has primarily been on the development and application of smart home technologies. However, comprehensive studies examining the general landscape of diverse research areas for smart home health care are still lacking.

Objective: This study aims to determine the intellectual structure of smart home health care in a time series by conducting a coword analysis and topic analysis. Specifically, it investigates (1) the intellectual basis of smart home health care through overall academic status, (2) the intellectual foci through influential keywords and their evolutions, and (3) intellectual trends through primary topics and their evolutions.

Methods: Analyses were conducted in 5 steps: (1) data retrieval from article databases (Web of Science, Scopus, and PubMed) and the initial dataset preparation of 6080 abstracts from the year 2000 to the first half of 2019; (2) data preprocessing and refinement extraction of 25,563 words; (3) a descriptive analysis of the overall academic status and period division (ie, 4 stages of 3-year blocks); (4) coword analysis based on word co-occurrence networks for the intellectual foci; and (5) topic analysis for the intellectual trends based on latent Dirichlet allocation (LDA) topic modeling, word-topic networks, and researcher workshops.

Results: First, regarding the intellectual basis of smart home health care, recent academic interest and predominant journals and research domains were verified. Second, to determine the intellectual foci, primary keywords were identified and classified according to the degree of their centrality values. Third, 5 themes pertaining to the topic evolution emerged: (1) the diversification of smart home health care research topics; (2) the shift from technology-oriented research to technological convergence research; (3) the expansion of application areas and system functionality of smart home health care; (4) the increased focus on system usability, such as service design and experiences; and (5) the recent adaptation of the latest technologies in health care. Based on these findings, the pattern of technology diffusion in smart home health care research was determined as the adaptation of technologies, the proliferation of application areas, and an extension into system design and service experiences.

Conclusions: The research findings provide academic and practical value in 3 aspects. First, they promote a comprehensive understanding of the smart home health care domain by identifying its multifaceted intellectual structure in a time series. Second, they can help clinicians discern the development and dispersion level of their respective disciplines. Third, the pattern of technology diffusion in smart home health care could help scholars comprehend current and future research trends and identify research opportunities based on upcoming research waves of newly adapted technologies in smart home health care.

(J Med Internet Res 2021;23(1):e19625) doi:10.2196/19625

KEYWORDS

smart home; smart home health care; coword analysis; topic analysis; intellectual structure; academic trends

Introduction

The rapid development of information and communication technologies, including the Internet of Things (IoT) and ubiquitous computing, have made home environments highly intelligent and allow smart homes to be realized [1]. In previous research, the smart home was characterized as the integration of technologies such as home automation, automatically controlled systems, communication networks and connecting devices and services, remote access and control, and home intelligence, with the context awareness of users [2,3]. Accordingly, the purpose of smart homes has been defined as providing a better home life experience with enhanced security, safety, communication, comfort, and entertainment through technical management of the home environment [4]. On the other hand, ethical and legal concerns surrounding smart home technologies have been pointed out regarding the security, privacy, and confidentiality of users [5,6]. Nonetheless, with a prudent approach to these ethical and legal challenges, smart homes could be an effective tool for continuous, remote, and nonintrusive health monitoring and disease prevention while guaranteeing users' independence and quality of life [7,8].

Various terms have been used to describe such technology, such as "smart homes in or for health care" [7,8], "health smart home" [9,10], and "ubiquitous health care" [11,12]. For consistent terminology, we adopted the term "smart home health care" and established the following operationalized definition [13], which embraces both technical and experiential perspectives:

> *Smart home health care is a health care service in one's residence incorporated with IoT technology and ubiquitous computing, which has the characteristics of home automation and home intelligence, communication networks, and remote access and control by authorized health care personnel. It offers informal health care services such as real-time or long-term health monitoring, unobtrusive activity support without interference with daily lives, and disease prevention through anomaly detection. It can reduce care costs, allow satisfactory service experience in a comfortable and private home environment, and ensure the independence of residents.*

The academic focus of previous smart home research has mainly been on the development and application of smart home technologies in the fields of computer science and engineering [2-4,14]. Likewise, smart homes for health care have primarily been examined from the perspective of technological application [7,8,15,16]. Such studies have analyzed and classified smart home health care services according to the type of sensor, network or communication technology, and algorithm models of data processing [3,7,8]. Despite the recent increase of technology-related research on smart home health care, comprehensive studies on the general state of its diverse research areas remain lacking.

Therefore, it is necessary to comprehend the current landscape of smart home health care research and seek future research opportunities. The specific research questions we used to investigate the intellectual structure of smart home health care are the following: (RQ1) What is the overall academic status of the current research on smart home health care, and which research fields have mainly been focused on? (RQ2) What are the representative keywords in the research on smart home health care, and how have they evolved over time? (RQ3) What are the intellectual trends in smart home health care research, and what are the main research orientations?

Thus, the objective of this study is to determine representative research topics on smart home health care and their evolutionary trends, and determine the intellectual structure of smart home health care research by conducting bibliometric network analysis. The resulting findings could lead to a comprehensive understanding of the current literature on smart home health care and enable scholars to extend their academic interests in future research.

Methods

Bibliometric Network Analysis and Coword Analysis

Bibliometric network analysis refers to a computer-assisted scientific review methodology that provides quantitative and statistical analysis by summarizing a large number of research publications through various descriptors and indicators [17]. Havemann and Scharnhorst [18] organized the approaches of bibliometric networks into the following categories: (1) citation networks of articles and journals; (2) bibliographic coupling networks; (3) cocitation networks of articles, authors, and journals; (4) co-authorship networks; and (5) word co-occurrence networks (ie, coword analysis). Among them, coword analysis has been considered an effective approach for understanding key topics in a certain research area, calculating the association strength of representative terms, and illustrating the field's knowledge structure by revealing patterns and trends among those topics [19,20]. Researchers have combined coword analysis with cluster analysis, multidimensional scaling, or social network analysis (SNA) to statistically investigate word co-occurrence patterns using titles, keywords, and abstracts [21,22]. In health informatics research, the combination of coword analysis and SNA has been adopted to identify the trends of specific themes in health care, such as mobile health [23], cybersecurity [24], and robotic or mixed reality surgery [25,26]. Accordingly, we adopted coword analysis combined with SNA to better understand key topics in smart home health care research. The analysis flow in this study was adopted from previous research [27,28] and included data retrieval, data preprocessing and refinement, and data analysis.

Topic Analysis Combining Topic Modeling and Social Network Analysis

To complement the coword analysis with word co-occurrence networks, we adopted topic modeling to identify more detailed topic groups and their evolutionary trends. In machine learning and natural language processing, topic modeling is a prominent technique for data mining, latent data detection, and finding associations among data and text documents to discover hidden semantic structures [29,30]. Latent Dirichlet allocation (LDA) [31], one of the most popular text mining methods [32], was adopted in this study to identify latent topics in the retrieved data and classify words and documents into topics. We also combined topic modeling with SNA to establish word-topic networks and better understand the relations among topics.

Data Collection and Analysis Process

Data Retrieval

We collected articles relevant to this research from the journal databases of Web of Science, Scopus, and PubMed. Web of Science and Scopus are two of the most authoritative scholarship databases, with peer-reviewed papers covering a wide range of subjects in various disciplines [33]. PubMed was also chosen in order to supplement domain-specific knowledge pertaining to health care. To retrieve the articles, a keyword search query was composed as follows:

Query = ((“intelligent” OR “smart”) AND “home”) AND ((“health” AND “care”) OR “ehealth” OR “mhealth” OR “uhealth”))

Article types were limited to journal articles and proceeding papers, article language was constrained to English, and the time period was limited to the year 2000 onwards, considering the recency of smart home technologies. The data retrieval was conducted on July 1, 2019, and 6579 articles were initially collected (4683 from Web of Science, 1421 from Scopus, and 475 from PubMed). From the raw dataset, we excluded data that were duplicated among databases or that had no information on the year of publication, author, or abstract. The initial dataset of 6080 articles was prepared using the following bibliometric data: title, year of publication, author, journal, abstract, and keywords.

Data Preprocessing and Refinement

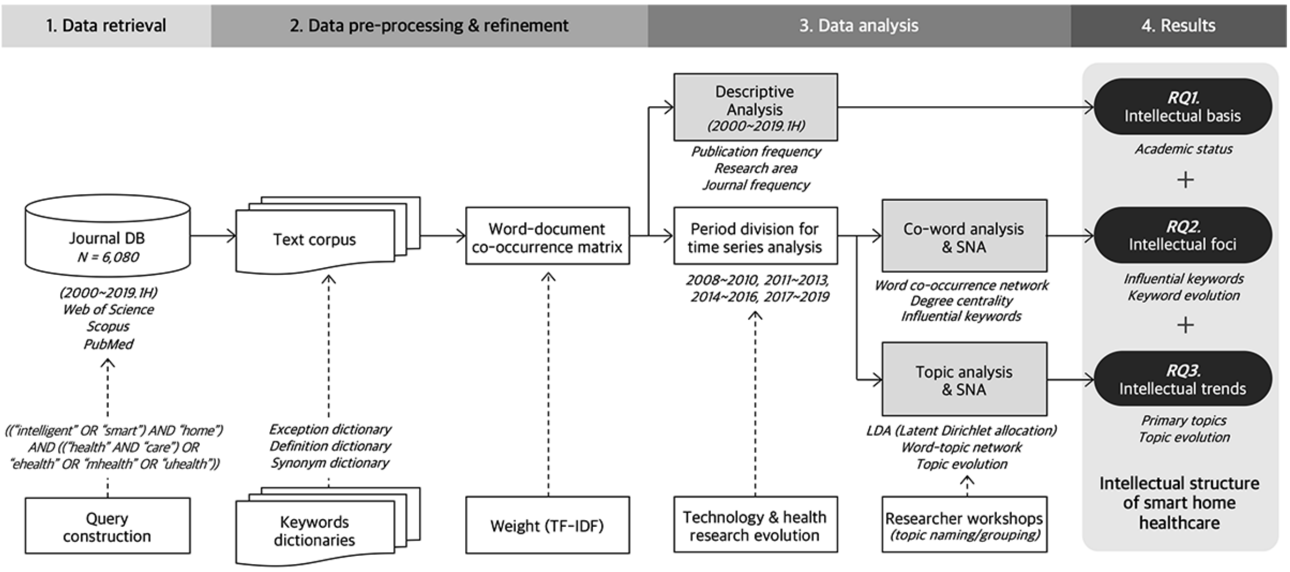
Noun data needed to be extracted from the abstract dataset in order to constitute a text corpus. For this, we conducted data preprocessing according to the stages of dictionary development: exception, definition, and synonyms. First, the “exception” dictionary was established to remove any general terminologies which are routinely utilized in abstracts to explain research structures, processes, and methodologies (eg, background, method, result, discussion, literature, purpose, implementation). Second, the “definition” dictionary was developed to define compound nouns as 1 word (eg, “smart home,” and not “smart” and “home” separately). To expand the definition dictionary, academic compound nouns were added to reflect the author keywords data. Third, the “synonym” dictionary was produced to detect duplicate meanings and replace them with 1 representative word (eg, “smart home” would represent “smart house,” “intelligent home,” “intelligent smart home,” and the plural forms of the same terms). The synonym dictionary also combined the singular and plural forms of the same terms into singular words. After this process, 25,563 words from 6080 abstracts were prepared as the text corpus for the following steps.

After defining the text corpus, data refinement was conducted in order to filter terms of high frequency but general usage. The dataset was arranged as a word-document matrix by occurrence, and was transformed into a co-occurrence matrix using the symmetric Ochiai-Salton algorithm. To filter normally used terminology, this study adopted the term frequency–inverse document frequency (TF-IDF) text-mining method. For data preprocessing and refinement, we utilized Python (version 3.7; Python Software Foundation [34]) and NetMiner (version 4.3; Cyram Inc [35]) commercial software. Since NetMiner software specializes in conducting semantic network analysis and offers a graphical demonstration of the network, it was also used in the following analysis stage to determine the networks and visualizations.

Data Analysis

Figure 1 depicts the overall process of this research methodology. The data analysis was conducted according to 3 subanalysis stages.

Figure 1. Research methodology. DB: database; RQ: research question; SNA: social network analysis; TF-IDF: term frequency–inverse document frequency.



First, we conducted a descriptive analysis based on the frequency of publication. From this, the research areas and principal journals relevant to smart home health care research could be identified. The results from this analysis could answer the first research question regarding overall academic status as an intellectual basis for smart home health care. After the descriptive analysis, time periods were divided by a certain scale in order to conduct further time-series analyses and explore various evolutions. To decide the time scale, we examined the evolution of technology and health research, or how health care research has evolved in response to recent emerging technologies. For this, we searched for several keywords directly pertinent to the study and emerging technology keywords [36] (eg, artificial intelligence [AI], IoT, and blockchain) associated with health care on the Web of Science database. Considering the evolution of yearly publication frequencies for those keywords, the time scale was divided into 4 stages of 3-year blocks: 2008-2010, 2011-2013, 2014-2016, and 2017-2019.

Second, we conducted a coword analysis combined with SNA to establish word co-occurrence networks. This was based on document frequency, TF-IDF values, and the degree centrality of keywords in word co-occurrence networks. Influential keywords and keyword evolution could then be determined, answering the second research question regarding the intellectual foci. We constructed word co-occurrence networks to transform the word-document networks into word-word networks, having weights of word co-occurrence frequency with a proximity measure of the correlation type inner product. Using degree-centrality analysis and link reduction, influential keywords were extracted by filtering the top 100-degree centrality words. Primary keywords (the top 40) were positioned according to their degree centrality values and compared within the time series to understand their evolution.

Third, we conducted a topic analysis combining topic modeling and SNA to establish word-topic networks. This allowed us to investigate the primary topics in smart home health care research and their evolutions within the time series, answering the third research question about intellectual trends. For topic modeling, we adopted an LDA with a Markov Chain Monte Carlo learning method to identify latent topics and classify words and documents into topics. Word-topic networks were then constructed and visualized according to allocation probability values. Three authors of this study held 3 consecutive researcher workshops to interpret the topic analysis results qualitatively to (1) name the individual topics from the results of LDA separately, (2) compare the name results of the 3 researchers and establish 1 set of topic names by reaching a consensus on any discrepancies, and (3) classify them into superordinate topic groups regarding the visualized word-topic networks. Lastly, the document frequencies allocated to the topic groups were counted and compared among different time periods to comprehend the topic evolution.

Results

Descriptive Analysis

Overall Academic Status

Using the initial 6080 articles, an overview of smart home health care research was compiled. Figure 2 shows the evolution of publication frequency from the year 2000, with the exponential curve being calculated to exclude the data on the first half of 2019. This shows that publication frequency rapidly increased in the last 10 years, illustrating the high level of research interest and academic popularity of smart home health care (R^2=0.979).

Figure 2. The evolution of publication frequencies from 2000-2019. (The publication frequency in 2019 was based on the first half of the year; thus, it was not included in calculating for the exponential curve.).

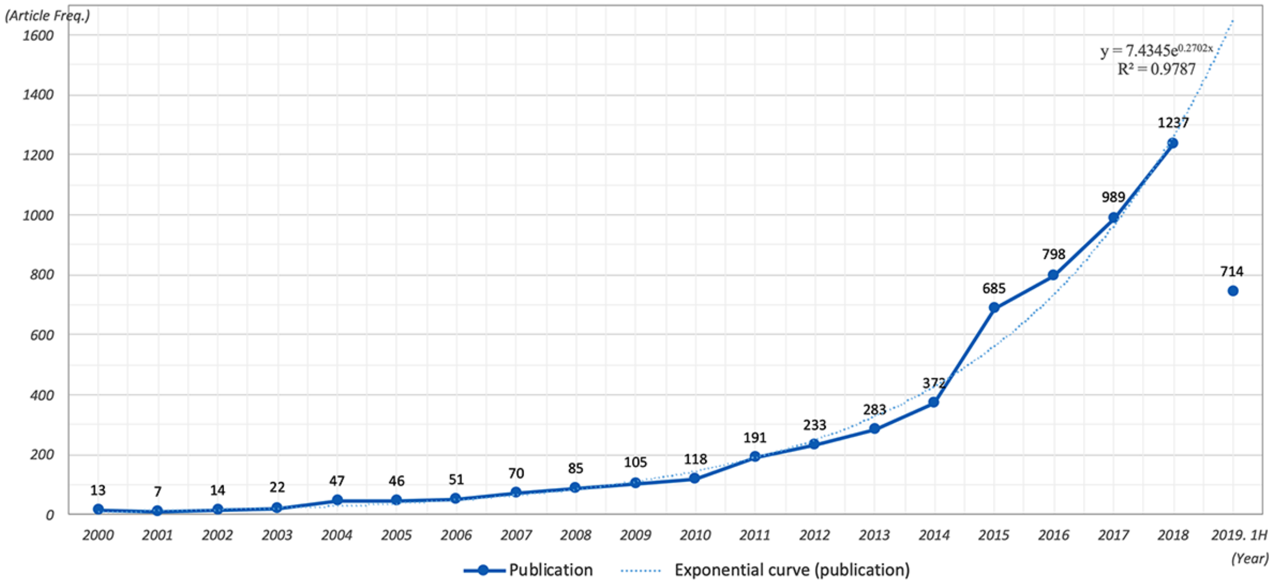


Table 1 exhibits the list of significant journals in smart home health care research from the year 2000; the journal frequencies of 1548 journals were counted. The *Journal of Medical Internet Research* holds the highest rank with 504 publications, followed by *JMIR mHealth and uHealth*, *JMIR Research Protocols*, *International Journal of Medical Informatics*, *BMC Medical*

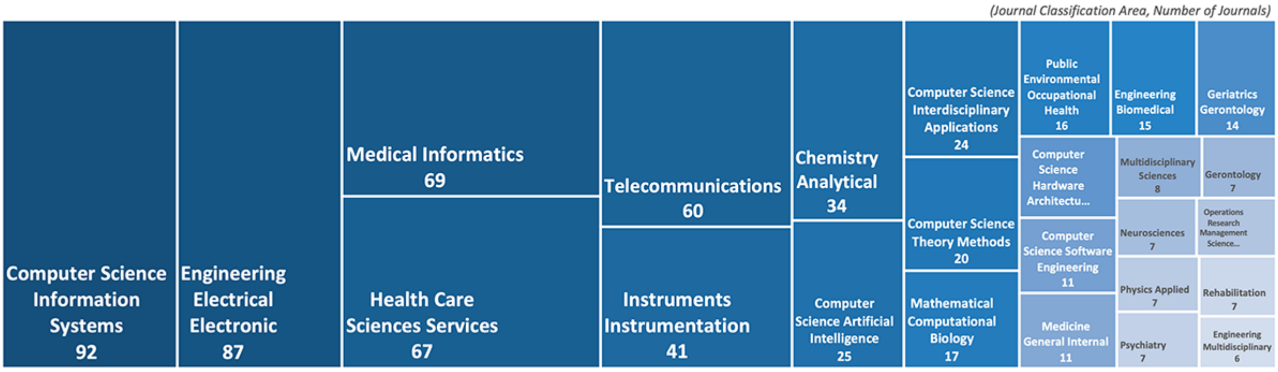
Informatics and Decision Making, *BMC Public Health,* and *JMIR Mental Health*. The top 3 journals from JMIR Publications constituted 17.62% (1071 of 6080 articles) of the total articles on smart home health care research.

Table 1. The top influential journals in smart home health care research from 2000-2019.

Journal	Article frequency
Journal of Medical Internet Research	504
JMIR mHealth and uHealth	392
JMIR Research Protocols	175
International Journal of Medical Informatics	91
BMC Medical Informatics and Decision Making	77
BMC Public Health	65
JMIR Mental Health	65
Journal of Medical Systems	63
Trials	59
Plos One	54
Journal of Telemedicine and Telecare	51
Studies of Health Technologies and Informatics	48
Methods of Information in Medicine	46
Journal of the American Medical Informatics Association	44
Sensors	44
BMJ Open	43
Health Informatics Journal	41
IEEE Journal of Biomedical and Health Informatics	33
IEEE Access	32
International Journal of Environmental Research and Public Health	32
Translational Behavioral Medicine	32
Journal of Health Communication	31
Patient Education and Counseling	31
ACM International Conference Proceeding Series	29
BMC Health Services Research	28
Telemedicine and eHealth	26
Contemporary Clinical Trials	25
JMIR Cancer	23
Aids and Behavior	22
CIN: Computers, Informatics, Nursing	22
Global Health Action	22
Journal of Biomedical Informatics	22

According to the academic area classification of Web of Science, the subject or type of journals that are pertinent to smart home health care research, derived with the same search query, is shown in Figure 3. The most predominant research domains are still from the engineering side, such as *Computer science information system* (23.77%) and *Electrical or electronic engineering* (22.48%). However, studies from medical and health care areas, such as *Medical informatics* (17.83%) and *Health care sciences and services* (17.31%), also constitute considerable portions of smart home health care research. Various medical fields, such as *Geriatrics or gerontology*, *Neurosciences*, *Rehabilitation*, and *Psychiatry*, emerged as small portions of the total research since advanced smart home health care technologies have already been heavily investigated in engineering domains.

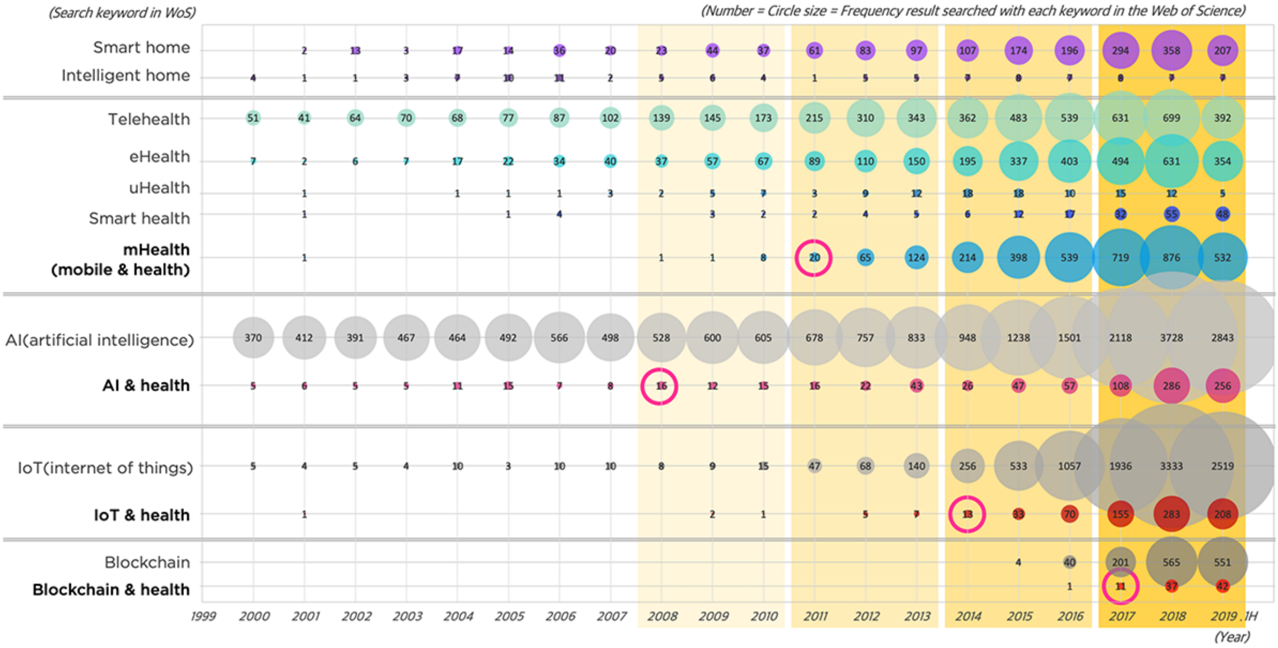
Figure 3. The academic area classification of journals on smart home health care research by Web of Science.



Period Division in Association with Emerging Technologies

As explained in the Methods section, the evolution of technology in association with health research was examined in order to select a time scale for further time-series analyses. The advent of new technologies can influence health care research and be adopted in health care and medical fields in diverse ways. Hence, several keywords used in the query of this study and emerging technology keywords [36] coupled with health care were searched in the Web of Science, and the publication frequencies were counted yearly from the year 2000. The result is illustrated in Figure 4.

Figure 4. The evolution of publications relevant to emerging technologies and health research from 2000-2019.



We specifically focused on the keywords “mHealth” (ie, mobile health), “AI and health,” “IoT and health,” and “blockchain and health,” which are technologies with significant influence on data collection and processing, service delivery, and interactions. The moments of advent were identified as the moments where publication frequency doubled compared to the previous year, exceeded 10 for the first time, and continued to gradually increase; this occurred for AI and health in 2008, mHealth in 2011, IoT and health in 2014, and blockchain and health in 2017. As AI had been greatly researched prior to 2000, the adaptation of AI in the health care field also occurred early, particularly in AI-assisted medical diagnosis research. Given this background, the advent year of the AI and health topic could be decided earlier than those of other topics. In accordance with the distribution of frequency, a time scale of 4 stages of 3-year blocks (2008-2010, 2011-2013, 2014-2016, and 2017-2019) was chosen. Despite the existence of telehealth research in the years between 2000 and 2007, this initial period was not included in the time-series analysis because the number of articles published at the time was insufficient to conduct a bibliometric network analysis, as shown in Figure 2.

Coword Analysis for the Intellectual Foci

In accordance with the time scale division, Table 2 shows general information on the words and documents per period. Over the total period, 25,556 words were extracted from 5810 documents. Approximately 75.0% of keywords occurred in 1-3 documents, 10.2% in 4-7 documents, and 5.0% in 8-13 documents. On average, any given keyword appeared in 9.2 documents, with the actual number of documents in which a given keyword appeared ranging from 1 to 813. The number of documents drastically increased throughout 4 periods, and period

4 contained half of the total documents (50.6%) despite the data in 2019 only comprising half a year. The number of words has also considerably grown, with the words from period 4 occupying 65.9% of the total 25,556 words.

Both the number of links created from word-document networks and those calculated by TF-IDF (which can represent how significant a word is in a certain document by eliminating ordinary words with high frequency but low importance) constituted more than half of the total networks. The evolution of documents and keywords could also imply that smart home health care research has been rapidly developing quantitatively and qualitatively.

Table 2. General information of the time-series analysis.

Variables	Total period (2008-2019)	Period 1 (2008-2010)	Period 2 (2011-2013)	Period 3 (2014-2016)	Period 4 (2017-2019)
Number of documents, n (%)	5810 (100)	308 (5.3)	707 (12.2)	1855 (31.9)	2940 (50.6)
Number of words, n (%)	25,556 (100)	3814 (14.9)	6868 (26.9)	12,749 (49.9)	16,831 (65.9)
Number of links in the word-document network, n (%)	333,921 (100)	14,724 (4.4)	36,642 (11.0)	107,405 (32.2)	175,150 (52.5)
Number of links in the word-document network, TF-IDF (%)	223,422 (100)	10,643 (4.8)	25,342 (11.3)	72,230 (32.3)	115,207 (51.6)

Among the word-document networks, the links whose TF-IDF values were above 0.5 and the top 10% of words in terms of document frequency were extracted. Then, word co-occurrence networks were also constructed in order to determine the primary keywords based on precise figures, namely degree centrality. Since the links from the original networks were too massive (as shown in Table 3), they were extracted in 2 steps: (1) by the top 100 words in degree centrality values, and (2) by link reduction according to a threshold of word co-occurrence frequency. "Network degree centralization" is a measurement that assesses the degree of inequality compared with a perfect "star network," which is the most unequal network type [37]. The values of the network degree centralization index changed during the time series, showing a relative decentralization in periods 2 and 4. This may indicate a horizontal expansion of research toward the significant subject areas during those periods.

Table 3. Properties of word co-occurrence networks by time series.

Properties	Period 1 (2008-2010)	Period 2 (2011-2013)	Period 3 (2014-2016)	Period 4 (2017-2019)
Word co-occurrence network information, n				
Links in the original network	184,550	414,721	1,084,449	1,576,645
Links in the extracted network 1 (by the top 100 words in degree centrality values)	4053	4705	4943	4931
Links in the extracted network 2 (by link reduction)	236	302	318	380
Threshold of co-occurrence frequencies for link reduction[a]	7	10	25	12
Network degree centralization, index %	30.2	23.7	47.1	34.4
Distribution of degree centrality				
Mean of degree centrality	0.048	0.061	0.064	0.077
Standard deviation	0.074	0.072	0.079	0.077
Min. of degree centrality	0.000	0.000	0.000	0.000
Max. of degree centrality	0.343	0.293	0.525	0.414

[a] The threshold value was decided by a link reduction simulation, identifying the point where the network component number suddenly increased.

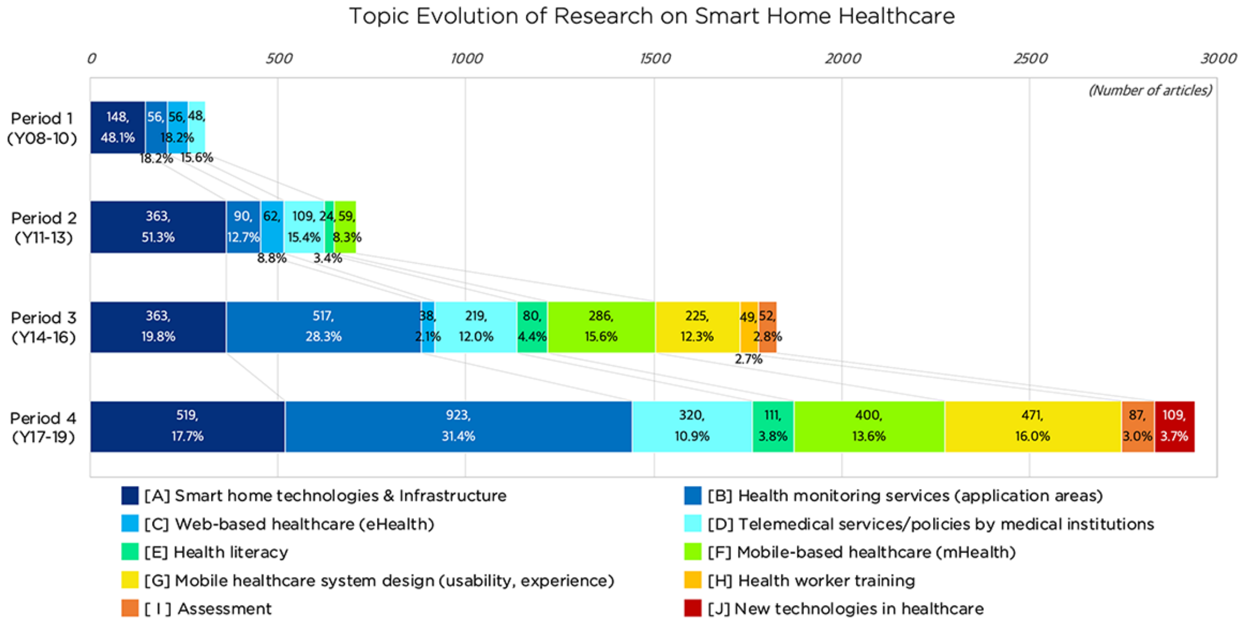
Moreover, the evolution of word co-occurrence networks was visualized using document frequency as the node size for each word, as exhibited in Multimedia Appendix 1. In period 1, the keywords "monitoring," "smart home," "communication," "home care," and "environment" were highlighted, while most of the remaining keywords were not sufficiently distinguished. In period 2, the top keywords in period 1 gained more prominence, and several additional words (eg, "sensor," "activity," and "work") noticeably emerged. In period 3, mHealth-relevant keywords, such as "mobile phone" and "SMS," became distinct. In period 4, existing keywords (eg, "mobile phone," "mobile app," "barrier," "experience," and "strategy") expanded their links. In comparison to prior periods, particularly period 1, the emphasized keywords in period 4 depicted a relatively even distribution of weights (ie, size of nodes); this implies that various research topics had been generated and investigated by many studies.

To determine the influential keywords and keyword evolutions in response to the second research question on intellectual foci, 40 primary keywords were determined based on their degree centrality values from the word co-occurrence networks (as shown in Multimedia Appendix 2). The primary keywords were categorized into the different groups based on their degree centrality values: core, semiperiphery, and periphery. The centrality value thresholds were decided based on the number of keywords in each group according to the data on degree centrality [38]. “Monitoring and (probably remote) communication in smart homes for home care of the elderly” was the main theme combining the core keywords in period 1. In period 2, the centrality values of “smart home,” “activity,” and “environment” increased in the core group, while “sensor” moved from the semiperiphery group to the core group. It may be inferred that research on smart home health care technologies, such as cases utilizing various types of sensors, was heavily conducted in this stage. In period 3, with the proliferation of mHealth research, keywords such as “mobile phone” and “SMS” shifted from the semiperiphery or periphery group to the core group. In period 4, “mobile app” transferred from the semiperiphery to the core group, while “barrier” and “experience” leaped from the periphery to the core group. This may imply that the applicability and user experience (eg, engagement, acceptability) of smart home health care services gained strong research attention in this period. The primary keywords extracted from the centrality analysis showed consistent results with the highlighted keywords in the previous word co-occurrence networks; this may enhance the validity of the deductive process in determining intellectual keywords and their changes.

Topic Analysis for the Intellectual Trends

As explained in the Methods section, a topic analysis combining topic modeling and SNA was conducted in order to investigate the primary research topics on smart home health care and their evolutions; this would address the third research question on intellectual trends. During the LDA topic modeling process, the number of topics were decided based on the number of documents in each period (ie, 13 topics in period 1, 20 topics in period 2, and 30 topics in both periods 3 and 4). According to the allocation probability of a word per topic, word-topic networks were constructed and visualized, extracting the top 100 to 150 words in probability values for effective plotting. During the 3 researcher workshop sessions, we established the name of each topic, embracing the top 5 keywords of allocation probability individually in the first session and together in the second, in order to arrive at a consensus regarding any discrepancies. In the last session, we classified the topics into groups, considering their meanings and network dispositions from the visualized word-topic networks. Figure 5 illustrates the topic evolution, showing the frequencies and percentages of topics in each period; Table 4 displays the topic analysis results of topic naming and grouping; and Multimedia Appendix 3 visualizes the word-topic networks associated with topic groups.

Figure 5. Topic evolution of research on smart home health care.



Through the analysis of topic modeling and word-topic networks, the dynamics of topic evolution were used to identify the intellectual trends regarding smart home health care. These trends can be summarized in 5 aspects: (1) the diversification of smart home health care research topics; (2) the shift from technology-oriented research to technological convergence research; (3) the expansion of application areas and system functionality of smart home health care; (4) the increased focus on system usability, such as service design and experiences; and (5) the recent adaptation of the latest technologies in health care.

Table 4. Topic analysis results from (1) LDA topic modeling, (2) word-topic network, (3) topic naming and grouping workshops.

Period and topic group	Topic number	Topic name	Document frequency
Period 1			
A	T03	Telemonitoring	20
	T04	Monitoring technology/infrastructure	24
	T05	IT-based decision support	18
	T09	Smart home activity detection	59
	T10	eHealth infrastructure	27
B	T01	Lifetime family care	15
	T06	Elderly care	13
	T08	Robot-based health service	15
	T13	Health event management	13
C	T07	Web-based health education	21
	T11	Web-based health contents delivery	35
D	T02	Security architecture in hospitals	27
	T12	Care delivery strategy	21
Period 2			
A	T01	Elderly care	66
	T02	Network integration	30
	T04	Signal/image detection	31
	T09	Communication platform	37
	T12	Privacy and security	40
	T13	Classification algorithm	25
	T14	Smart home activity detection	79
	T15	Monitoring technology/infrastructure	55
B	T03	Chronological disease	26
	T05	Clinical service	20
	T06	PHR based on social media	22
	T20	Health activity management	22
C	T07	Online patient community	41
	T17	Age-sensitive web accessibility	21
D	T08	Telemedical service	33
	T11	Primary care delivery	26
	T19	Care delivery policy	50
E	T10	eHealth literacy	24
F	T16	Healthcare ICT investment (mobile)	38
	T18	Mobile-based health contents delivery	21
Period 3			
A	T01	IoT and sensor network system	146
	T04	Activity sensing algorithm	126
	T13	Telemonitoring infrastructure	44
	T18	Privacy and security	47
B	T06	e-Fitness	24
	T08	Chronological disease	46

Period and topic group	Topic number	Topic name	Document frequency
	T12	Self-management	54
	T15	Exercise management	106
	T16	Elderly care	105
	T20	Cancer management	45
	T22	Mental health management	58
	T26	Lifetime family care by gamification	29
	T27	Chronological disease management	50
C	T11	eHealth service benefit	38
D	T07	Telemedical service	40
	T28	Care delivery policy	179
E	T24	Health literacy	80
F	T02	Health behavior awareness (mobile)	24
	T10	Age/gender-sensitive notification	58
	T17	Medication management (mobile)	80
	T25	Community healthcare (mobile)	72
	T29	mHealth data communication	52
G	T05	mHealth technology acceptance	30
	T14	eHealth system design	63
	T21	mHealth app design	56
	T23	Usability for health workers/patients	49
	T30	System design	27
H	T09	Health worker training	49
I	T03	Assessment accuracy	52
N/A[a]	T19	Etc	26
Period 4			
A	T19	Activity sensing algorithm	212
	T20	Smart home technology infrastructure	307
B	T03	Activity management	120
	T12	Chronological disease management	67
	T13	Mental healthcare	73
	T17	Emergency management	72
	T18	Rehabilitation training	70
	T21	Cancer management	89
	T22	Mental healthcare	130
	T24	Chronological disease management	56
	T25	Telemonitoring	71
	T26	Weight management	95
	T28	Age/gender-sensitive service	80
C	N/A	N/A	N/A
D	T02	Care cost-effectiveness (hospital)	61
	T05	Telemedicine platform	78
	T10	Community service	181

Period and topic group	Topic number	Topic name	Document frequency
E	T04	Health literacy	111
F	T01	Mobile support for smoking cessation	109
	T07	Lifetime family care (mobile)	88
	T15	Healthcare service providing (mobile)	117
	T23	Medication management (mobile)	86
G	T06	mHealth system design	160
	T14	Health system design	37
	T16	Elderly technology acceptance	69
	T27	Medical healthcare app	65
	T29	Technology acceptance	75
	T30	Mobile health system design	65
H	N/A	N/A	N/A
I	T11	Assessment algorithm	87
J	T08	Blockchain and healthcare	66
	T09	Telemedicine/telesurgery technology	43

[a] N/A: not applicable.

The research topics on smart home health care were diverse throughout the 4 time periods. Based on the preliminary topics from LDA modeling, 10 topic groups were identified: (A) smart home technologies and infrastructure, (B) health monitoring services (application areas), (C) web-based health care (eHealth), (D) telemedical services and policies by medical institutions, (E) health literacy, (F) mobile-based health care (mHealth), (G) mobile health care system design (usability, experience), (H) health worker training, (I) assessment, and (J) new technologies in health care

It was found that the number of topic groups present increased in every period. As shown in Table 4, period 1 started with 4 topic groups, namely, "smart home technologies and infrastructure," "application areas of health monitoring services," "web-based health care (eHealth)," and "telemedical services and policies by medical institutions." In period 2, 2 topic groups—"health literacy" and "mobile-based health care (mHealth)"—emerged. In period 3, another 3 topic groups were added: "mobile health care system design," "health worker training," and "assessment." Notably, in period 4, one topic group, "new technologies in health care," was added, and 2 topic groups, "web-based health care (eHealth)" and "health worker training," were withdrawn. Though the topic group on eHealth decreased continuously from period 1, this does not mean that research on eHealth simply vanished; the relevance of eHealth to smart home health care could have simply reduced when compared to that of mHealth, as represented by topic groups (F) and (G).

Second, results showed that the majority of the research focus gradually shifted from technology-oriented research to technological convergence research. In Table 4, in topic group A, diverse smart home health care technology-oriented research topics were investigated, such as "smart home activity detection" (P1-T09, P2-T14), "monitoring technology/infrastructure" (P1-T04, P2-T15), "privacy and security issues in smart home for health care technology" (P2-T12, P3-T18), "IoT and sensor network system" (P3-T01), "activity sensing algorithm" (P3-T04, P4-T19), and "smart home technology or infrastructure (P4-T20)." The portion of this topic group constituted approximately 50% of all articles in periods 1 and 2, but it shrank to below 20% in periods 3 and 4, even though their frequencies slightly increased. Instead, technological convergence was enlarged in smart home health care research, investigating diverse application areas of smart home health care services, as explained in the next aspect.

Third, the application areas and functionality of smart home health care expanded as topic group B (Table 4) attained the highest frequency in periods 3 and 4. The subtopics in this group showed diverse areas of smart home health care, such as "mental health management" (P3-T22, P4-T13/T22), "chronological disease management" (P2-T03, P3-T08/T27, P4-T12/T24), "activity management" (P2-T20, P4-T03), "cancer management" (P3-T20, P4-T21), "elderly care" (P1-T06, P3-T16), "exercise management" (P3-T15), "weight management" (P4-T26), "emergency management" (P4-T17), and "rehabilitation training" (P4-T18), in their order of frequency. Moreover, in the early periods, the application areas were more focused on extended medical care, such as the management of elderly care and chronological disease, or "clinical services" (P2-T05). However, in the later periods, diverse health care services in daily lives, such as "lifetime care by gamification" (P3-T26) and "e-fitness" (P3-T06), emerged.

Fourth, the focus on system usability, such as service design experiences, increased from period 3 (as demonstrated by topic group G, Table 4). The major subtopics with high frequencies in this group were the aspects of system or application design issues, such as "mHealth system design" (P4-T06, T30), "mHealth app design" (P3-T21, P4-T27), and "eHealth system design” (P3-T14). In addition, the user aspects of service experiences were also accentuated in periods 3 and 4, as shown by “technology acceptance” (P4-T29), in association with the unified theory of acceptance and use of technology [39,40], and “usability for health workers or patients” (P3-T23) [41,42].

Lastly, the results showed that the latest technologies were being adapted in smart home health care research, as seen in topic group J (Table 4), which emerged in period 4. The subtopics of this group were “blockchain and health care” (P4-T08) and “telemedicine or telesurgery technology” (P4-T09). As illustrated in Figure 4, blockchain technology emerged just before period 4, and research on the topic coupled with health care began to increase in association with data privacy and security [43,44] and system interoperability [45]. Furthermore, the role of telemedicine and smart home health care services in surgery domains, as part of perioperative or postoperative procedures, gained more academic interest during this period [46].

Discussion

Principal Findings

Based on the analysis results, several discussion points emerged regarding the intellectual structure of smart home health care research.

First, mobile technologies have broadened the scope of private spaces and accelerated the expansion of home care to smart health care services. In the past, personal health care was associated with the home environment, but the intervention of mobile or smartphones has expanded the boundary of personal health care beyond the home. Accordingly, the adoption of mobile technologies in smart home services has increased and diversified the service areas, service functions, usability, and experiences of home care services. As shown in Multimedia Appendix 2, “mobile phone” and “mobile app” emerged in the core group in periods 3 and 4, instead of “smart home,” which was in the core in periods 1 and 2. This may imply that the home concept includes mobile personal space with regard to health care.

Second, smart technology evolution has enabled hospital-led services to be offered as personalized services. Hospital-led telehealth research in the past has been infused with home health care research, and the mobile technologies that are relevant to physiological monitoring and privacy protection have enhanced the association of hospital care and personal home care. While a large part of existing research in medical informatics used to focus on electronic medical records (EMR), recent research interest has been focused on personal health records (PHR) due to the adaptation of enabling mobile technologies (eg, IoT, sensor technology, blockchain) [47,48]. Moreover, mobile technologies could realize the concept of connected hospital or extended medical experiences, enhancing remote medical examination, treatment, and management.

Third, the pattern of technology innovation diffusion [49,50] in smart home health care research may be identified from the topic evolution results. Based on the findings in the topic evolution analysis, the technology diffusion in smart home health care fields has evolved from (1) the adaptation of technologies and (2) proliferation of application areas to (3) extension into system design and service experiences. Thus, it may be anticipated that the perspective of service design and experiences with diverse smart home health care application areas would be developed in the near future. Furthermore, when new technology appears in health care industries, a similar pattern of research evolution could be expected. For instance, blockchain technologies, which were recently adapted in smart home health care research, became an emerging topic (Table 4, P4-T08) in period 4. This technology could lead to a further investigation of the diverse application areas of smart home health care, resulting in a smart home health care system design coupled with blockchain technology and user experience features.

Lastly, this pattern of the technology diffusion process in smart home health care could have particular aspects caused by the medical or health care field's unique characteristics. Ethical and legal issues pertaining to personal health data are frequent concerns when technologies are integrated into health care services. Policies, regulations, and legislation systems relating to privacy and security issues in smart home health care services require greater attention than other commercial services [51,52]. Moreover, health care services usually demand a large number of clinical trials and verifications in order to generalize their applicability. Therefore, the technology diffusion process in smart home health care may take more prudence and time to evolve into further stages than other types of smart services. Nevertheless, as shown in Figure 4, the time lag of the adaptation of state-of-the-art technology in health care industries becomes shorter (eg, from AI to “AI and health,” from IoT to “IoT and health,” and from blockchain to “blockchain and health”). With the fourth industrial revolution, the cycle of technology diffusion in smart home health care could be gradually reduced.

Limitations and Further Research

Despite the findings on the intellectual structure of smart home health care research, this study has several limitations due to its methodological approach. First, as coword analysis and topic analysis are quantitative and statistical analyses that summarize a large number of publications, qualitative research would be required in order to investigate specific topics in-depth among the various topic groups exhibited in Table 4.

Coword analysis is suggested as a prior stage to a systematic review to guide and accelerate the review process [53]. Therefore, a qualitative and thorough systematic literature review investigating specific topic groups can be conducted in future research. Second, this study adopted a semantic network analysis approach based on keywords and abstracts in order to identify keywords and topics in smart home health care research. However, the traditional approach in bibliometric analysis, which is based on co-authorship and citation information, could be applicable in future studies to investigate knowledge flows, research groups, and organizational influences. Third, this study extracted a text corpus only from noun data, following the general approach of semantic network analysis. However, if a text corpus is extracted from both adjectives and nouns, a more specific context in the usage of noun data may be explored with further analysis. Last, the text corpus was established from academic articles; thus, the perspective of commercial and business-wise cutting-edge technologies in the health care industry might not have been sufficiently included in this study's findings.

Conclusions

This research aimed to determine the representative research keywords and topics in smart home health care research and their evolutionary trends to demonstrate the intellectual structure of smart home health care by conducting bibliometric network analysis. First, we identified the academic status of the intellectual basis for smart home health care. With the recent increase in academic interest, smart home health care has been primarily investigated in research domains such as computer science information systems, electrical engineering, medical informatics, and health care science and services. Second, keywords and centrality analysis were conducted based on word co-occurrence networks to determine the intellectual foci. The evolution of primary keywords was demonstrated throughout 4 periods, and the top 3 keywords in each period were identified: "monitoring, communication, smart home," "smart home, monitoring, sensor," "mobile phone, SMS, monitoring," and "mobile app, barrier, experience," respectively. Third, we conducted a topic analysis based on topic modeling and word-topic networks in order to determine the intellectual trends. Accordingly, 5 aspects of topic evolution emerged: (1) the diversification of smart home health care research topics, (2) the shift from technology-oriented research to technological convergence research, (3) the expansion of application areas and system functionality of smart home health care, (4) the increased focus on system usability, such as service design and experiences, and (5) the recent adaptation of the latest technologies in health care.

Through these findings, this study discussed the role of mobile and smart technologies in relation to the expansion of home care to smart health care services, and the extension of hospital-led services to personalized services. Moreover, the pattern of technology diffusion in smart home health care research was verified as (1) the adaptation of technologies, (2) the proliferation of application areas, and then (3) the extension into system design and service experiences. This pattern may take more time to evolve due to the ethical and legal concerns surrounding personal health data; however, the cycle of technology diffusion in smart home health care is expected to shorten gradually.

The value and potential contributions of these findings can be summarized in 3 aspects. First, they promote a comprehensive understanding of the smart home health care domain by determining its intellectual structure in a time series. The results disclosed and structuralized the massive knowledge body of smart home health care, which has been investigated from diverse research domains but still lacked in comprehending the general landscape. Second, these findings can help clinicians recognize the development and dispersion level of their disciplines and specialty topics. Moreover, practitioners can recognize other related topics to their major disciplines and seek any cooperation opportunities associated with smart home health care services. Third, the pattern of technology diffusion in smart home health care can enable scholars to understand current and future research trends. Particularly, this study is valuable because it confirmed that the pattern of technology diffusion could be applicable in smart home health care; thus, researchers can anticipate the upcoming research waves of newly adapted technologies in smart home health care and explore further research opportunities. Moreover, as a practical implication, the technological trends and their diffusion patterns in smart home health care can be utilized when professionals investigate market status, analyze their competitive advantages, and establish business strategies in their respective practice areas.

Acknowledgments

This work was supported by the Sungshin Women's University Research Grant of 2018-2-82-008. This fund has no specific role or influence in the research process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Word co-occurrence network evolution of research on smart home health care.
[PNG File , 1420 KB - jmir_v23i1e19625_app1.png]

Multimedia Appendix 2
The evolution of the position of the top 40 keywords by degree centrality in word co-occurrence networks.
[DOCX File , 318 KB - jmir_v23i1e19625_app2.docx]

Multimedia Appendix 3
Word-topic network evolution of research on smart home health care.
[PNG File , 1981 KB - jmir_v23i1e19625_app3.png]

References

1. Bassi A, Horn G. Internet of Things in 2020: Roadmap for the Future. European Commission: Information Society and Media. 2008 May 07. URL: https://docbox.etsi.org/erm/Open/CERP%2020080609-10/Internet-of-Things_in_2020_EC-EPoSS_Workshop_Report_2008_v1-1.pdf [accessed 2020-03-10]
2. Park SH, Won SH, Lee JB, Kim SW. Smart home: digitally engineered domestic life. Personal Ubiquitous Comput. (3-4) 2003;7:189-196. [doi: 10.1007/s00779-003-0228-9]
3. Alam MR, Reaz MBI, Ali MAM. A Review of Smart Homes—Past, Present, and Future. IEEE Trans. Syst., Man, Cybern. C 2012 Nov;42(6):1190-1203. [doi: 10.1109/TSMCC.2012.2189204]
4. Aldrich F. Smart homes: past, presentfuture. In: Harper R, editor. Inside the smart home. London: Springer Science & Business Media; 2006:17-40.
5. Pisani AR, Wyman PA, Mohr DC, Perrino T, Gallo C, Villamar J, et al. Human Subjects Protection and Technology in Prevention Science: Selected Opportunities and Challenges. Prev Sci 2016 Aug;17(6):765-778 [FREE Full text] [doi: 10.1007/s11121-016-0664-1] [Medline: 27220838]
6. Kelly P, Marshall SJ, Badland H, Kerr J, Oliver M, Doherty AR, et al. An ethical framework for automated, wearable cameras in health behavior research. Am J Prev Med 2013 Mar;44(3):314-319. [doi: 10.1016/j.amepre.2012.11.006] [Medline: 23415131]
7. Amiribesheli M, Benmansour A, Bouchachia A. A review of smart homes in healthcare. J Ambient Intell Human Comput 2015 Mar 14;6(4):495-517. [doi: 10.1007/s12652-015-0270-2]
8. Majumder S, Aghayi E, Noferesti M, Memarzadeh-Tehran H, Mondal T, Pang Z, et al. Smart Homes for Elderly Healthcare-Recent Advances and Research Challenges. Sensors (Basel) 2017 Oct 31;17(11):-1 [FREE Full text] [doi: 10.3390/s17112496] [Medline: 29088123]
9. Rialle V, Duchene F, Noury N, Bajolle L, Demongeot J. Health "Smart" home: information technology for patients at home. Telemed J E Health 2002;8(4):395-409. [doi: 10.1089/15305620260507530] [Medline: 12626109]
10. Bartolomeu P, Fonseca J, Vasques F. Challenges in Health Smart Homes. 2008 Presented at: Second International Conference on Pervasive Computing Technologies for Health care; Jan. 30 -Feb. 1; Jan 30-Feb 1, 2008; Tampere, Finland p. 19-22 URL: https://ieeexplore.ieee.org/document/4571016 [doi: 10.1109/pcthealth.2008.4571016]
11. Deen MJ. Information and communications technologies for elderly ubiquitous healthcare in a smart home. Pers Ubiquit Comput 2015 Jun 26;19(3-4):573-599. [doi: 10.1007/s00779-015-0856-x]
12. Rahmani A, Thanigaivelan N, Tuan NG, Granados J, Negash B, Liljeberg P, et al. Smart e-Health Gateway: Bringing intelligence to Internet-of-Things based ubiquitous healthcare systems. 2015 Presented at: 12th Annual IEEE Consumer Communications and Networking Conference (CCNC); 9-12 Jan. 2015; Las Vegas, NV, USA p. 826-834 URL: https://ieeexplore.ieee.org/document/7158084 [doi: 10.1109/ccnc.2015.7158084]
13. Kang H, Han J, Kwon GH. An Ecological Approach to Smart Homes for Health Care Services: Conceptual Framework of a Smart Servicescape Wheel. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e12425 [FREE Full text] [doi: 10.2196/12425] [Medline: 30735145]
14. Saad al-sumaiti A, Ahmed MH, Salama MMA. Smart Home Activities: A Literature Review. Electric Power Components and Systems 2014 Feb 05;42(3-4):294-305. [doi: 10.1080/15325008.2013.832439]
15. Chan M, Estève D, Escriba C, Campo E. A review of smart homes- present state and future challenges. Comput Methods Programs Biomed 2008 Jul;91(1):55-81. [doi: 10.1016/j.cmpb.2008.02.001] [Medline: 18367286]
16. Demiris G, Hensel BK. Technologies for an Aging Society: A Systematic Review of “Smart Home” Applications. Yearb Med Inform 2018 Mar 07;17(01):33-40. [doi: 10.1055/s-0038-1638580]
17. De Bellis N. Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. Maryland, USA: Scarecrow Press; 2009.
18. Havemann F, Scharnhorst A. Bibliometric Networks. 2012. URL: https://arxiv.org/abs/1212.5211 [accessed 2020-02-17]
19. Lee WH. How to identify emerging research fields using scientometrics: An example in the field of Information Security. Scientometrics 2008 Jul 12;76(3):503-525. [doi: 10.1007/s11192-007-1898-2]
20. Leydesdorff L, Welbers K. The semantic mapping of words and co-words in contexts. Journal of Informetrics 2011 Jul;5(3):469-475. [doi: 10.1016/j.joi.2011.01.008]
21. Hu J, Zhang Y. Research patterns and trends of Recommendation System in China using co-word analysis. Information Processing & Management 2015 Jul;51(4):329-339. [doi: 10.1016/j.ipm.2015.02.002]
22. Schrock WA, Zhao Y, Richards KA, Hughes DE, Amin MS. On the nature of international sales and sales management research: a social network–analytic perspective. Journal of Personal Selling & Sales Management 2018 Feb 20;38(1):56-77. [doi: 10.1080/08853134.2018.1428493]
23. Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z. Theme Trends and Knowledge Structure on Mobile Health Apps: Bibliometric Analysis. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18212 [FREE Full text] [doi: 10.2196/18212] [Medline: 32716312]
24. Jalali MS, Razak S, Gordon W, Perakslis E, Madnick S. Health Care and Cybersecurity: Bibliometric Analysis of the Literature. J Med Internet Res 2019 Feb 15;21(2):e12644 [FREE Full text] [doi: 10.2196/12644] [Medline: 30767908]

25. Han J, Kang H, Kim M, Kwon GH. Mapping the intellectual structure of research on surgery with mixed reality: Bibliometric network analysis (2000-2019). J Biomed Inform 2020 Sep;109:103516. [doi: 10.1016/j.jbi.2020.103516] [Medline: 32736125]
26. Shen L, Wang S, Dai W, Zhang Z. Detecting the Interdisciplinary Nature and Topic Hotspots of Robotics in Surgery: Social Network Analysis and Bibliometric Study. J Med Internet Res 2019 Mar 26;21(3):e12625 [FREE Full text] [doi: 10.2196/12625] [Medline: 30912752]
27. Heradio R, de la Torre L, Galan D, Cabrerizo FJ, Herrera-Viedma E, Dormido S. Virtual and remote labs in education: A bibliometric analysis. Computers & Education 2016 Jul;98:14-38. [doi: 10.1016/j.compedu.2016.03.010]
28. Du HS, Ke X, Chu SK, Chan LT. A bibliometric analysis of emergency management using information systems (2000-2016). OIR 2017 Aug 14;41(4):454-470. [doi: 10.1108/oir-05-2017-0142]
29. Blei DM. Probabilistic topic models. Commun. ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
30. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci 1990 Sep;41(6):391-407. [doi: 10.1002/(sici)1097-4571(199009)41:6<391::aid-asi1>3.0.co;2-9]
31. Blei D, Ng A, Jordan M. Latent Dirichllocation Allocation. The Journal of Machine Learning Research 2003;3:993-1022 [FREE Full text] [doi: 10.5555/944919.944937]
32. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2018 Nov 28;78(11):15169-15211. [doi: 10.1007/s11042-018-6894-4]
33. Meho LI, Yang K. Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. J. Am. Soc. Inf. Sci 2007 Nov;58(13):2105-2125. [doi: 10.1002/asi.20677]
34. Guzdial M, Ericson B. Introduction to Computing and Programming in Python, 4th Edition. Essex, England: Pearson; 2015.
35. Cyram I. NetMiner. Social Network Analysis Software. 2013. URL: http://www.netminer.com/main/main-read.do [accessed 2019-08-16]
36. Cearley D, Burke B, Searle S, Walker M. Top 10 Strategic Technology Trends for 2018. GartnerTrends. 2017. URL: http://brilliantdude.com/solves/content/GartnerTrends2018.pdf [accessed 2019-08-25]
37. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press; 1994.
38. Ronda-Pupo GA, Guerras-Martin L. Dynamics of the evolution of the strategy concept 1962-2008: a co-word analysis. Strat. Mgmt. J 2011 Aug 16;33(2):162-188. [doi: 10.1002/smj.948]
39. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
40. Tavares J, Oliveira T. Electronic Health Record Portal Adoption: a cross country analysis. BMC Med Inform Decis Mak 2017 Jul 05;17(1):97 [FREE Full text] [doi: 10.1186/s12911-017-0482-9] [Medline: 28679423]
41. Cruz Zapata B, Hernández Niñirola A, Idri A, Fernández-Alemán JL, Toval A. Mobile PHRs compliance with Android and iOS usability guidelines. J Med Syst 2014 Aug 24;38(8):81. [doi: 10.1007/s10916-014-0081-6] [Medline: 24957397]
42. Georgsson M, Staggers N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. J Biomed Inform 2016 Feb;59:115-129 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.008] [Medline: 26639894]
43. Ichikawa D, Kashiyama M, Ueno T. Tamper-Resistant Mobile Health Using Blockchain Technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
44. Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T. Secure and Scalable mHealth Data Management Using Blockchain Combined With Client Hashchain: System Design and Validation. J Med Internet Res 2019 May 16;21(5):e13385 [FREE Full text] [doi: 10.2196/13385] [Medline: 31099337]
45. Dimitrov DV. Blockchain Applications for Healthcare Data Management. Healthc Inform Res 2019;25(1):51. [doi: 10.4258/hir.2019.25.1.51]
46. Coldebella B, Armfield NR, Bambling M, Hansen J, Edirippulige S. The use of telemedicine for delivering healthcare to bariatric surgery patients: A literature review. J Telemed Telecare 2018 Oct 22;24(10):651-660. [doi: 10.1177/1357633x18795356]
47. Cunningham SG, Brillante M, Allardice B, Conway N, McAlpine RR, Wake DJ. My Diabetes My Way: supporting online diabetes self-management: progress and analysis from 2016. Biomed Eng Online 2019 Feb 08;18(1):13 [FREE Full text] [doi: 10.1186/s12938-019-0635-4] [Medline: 30736798]
48. Moon EW, Tan NC, Allen JC, Jafar TH. The Use of Wireless, Smartphone App-Assisted Home Blood Pressure Monitoring Among Hypertensive Patients in Singapore: Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 May 28;7(5):e13153 [FREE Full text] [doi: 10.2196/13153] [Medline: 30905872]
49. Peres R, Muller E, Mahajan V. Innovation diffusion and new product growth models: A critical review and research directions. International Journal of Research in Marketing 2010 Jun;27(2):91-106. [doi: 10.1016/j.ijresmar.2009.12.012]
50. Rogers E. Diffusion of Innovations, 5th Edition. NY, USA: Simon and Schuster; 2003.
51. Carter A, Liddle J, Hall W, Chenery H. Mobile Phones in Research and Treatment: Ethical Guidelines and Future Directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]
52. Nebeker C, Lagare T, Takemoto M, Lewars B, Crist K, Bloss CS, et al. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. Transl Behav Med 2016 Dec;6(4):577-586 [FREE Full text] [doi: 10.1007/s13142-016-0426-4] [Medline: 27688250]

53. Radhakrishnan S, Erbis S, Isaacs JA, Kamarthi S. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. PLoS ONE 2017 Mar 22;12(3):e0172778. [doi: 10.1371/journal.pone.0172778]

Abbreviations

AI: artificial intelligence
eHealth: web-based health
IoT: Internet of Things
LDA: latent Dirichlet allocation
mHealth: mobile health
SNA: social network analysis
TD-IDF: term frequency–inverse document frequency
uHealth: ubiquitous health (or health services delivered using ubiquitous technology, such as radio frequency identification)

Edited by R Kukafka; submitted 02.05.20; peer-reviewed by L Hess, L Moreira, M Behzadifar; comments to author 25.05.20; revised version received 29.09.20; accepted 10.01.21; published 21.01.21.

Please cite as:
Kang HJ, Han J, Kwon GH
Determining the Intellectual Structure and Academic Trends of Smart Home Health Care Research: Coword and Topic Analyses
J Med Internet Res 2021;23(1):e19625
URL: http://www.jmir.org/2021/1/e19625/
doi:10.2196/19625
PMID:33475514

Review

Understanding eHealth Cognitive Behavioral Therapy Targeting Substance Use: Realist Review

Farhud Shams[1], BSc; James S H Wong[1], BSc; Mohammadali Nikoo[1], MD; Ava Outadi[1,2], BA; Ehsan Moazen-Zadeh[3], MD, MSc; Mostafa M Kamel[1,4], MD; Michael Jae Song[5], BSc, MPH; Kerry L Jang[6], PhD; Reinhard Michael Krausz[1,7,8], MD, PhD, FRCPC

[1]Institute of Mental Health, Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

[2]Department of Educational and Counselling Psychology, and Special Education, Faculty of Education, University of British Columbia, Vancouver, BC, Canada

[3]Addiction Institute of Mount Sinai, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York City, NY, United States

[4]Department of Psychiatry, Tanta University, Tanta, Egypt

[5]Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

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

[7]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[8]Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Vancouver, BC, Canada

Corresponding Author:
Farhud Shams, BSc
Institute of Mental Health
Department of Psychiatry
University of British Columbia
430-5950 University Boulevard
David Strangway Bldg
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 6478345370
Email: f.shams@hotmail.de

Abstract

Background: There is a growing body of evidence regarding eHealth interventions that target substance use disorders. Development and funding decisions in this area have been challenging, due to a lack of understanding of what parts of an intervention work in which context.

Objective: We conducted a realist review of the literature on electronic cognitive behavioral therapy (eCBT) programs for substance use with the goal of answering the following realist question: “How do different eCBT interventions for substance use interact with different contexts to produce certain outcomes?”

Methods: A literature search of published and gray literature on eHealth programs targeting substance use was conducted. After data extraction, in order to conduct a feasible realist review in a timely manner, the scope had to be refined further and, ultimately, only included literature focusing on eCBT programs targeting substance use. We synthesized the available evidence from the literature into Context-Mechanism-Outcome configurations (CMOcs) in order to better understand when and how programs work.

Results: A total of 54 papers reporting on 24 programs were reviewed. Our final results identified eight CMOcs from five unique programs that met criteria for relevance and rigor.

Conclusions: Five strategies that may be applied to future eCBT programs for substance use are discussed; these strategies may contribute to a better understanding of mechanisms and, ultimately, may help design more effective solutions in the future. Future research on eCBT programs should try to understand the mechanisms of program strategies and how they lead to outcomes in different contexts.

(J Med Internet Res 2021;23(1):e20557) doi:10.2196/20557

KEYWORDS

eHealth CBT; substance use; realist review; eHealth mechanisms; opioid crisis

Introduction

An estimated 269 million individuals worldwide used drugs in 2018, which represents a 30% increase from 2009 [1]. Globally, around 36 million individuals have a drug use disorder; however, only 1 out of 8 individuals who need substance use treatment receive it [1]. In particular, North America is currently experiencing an unprecedented epidemic of drug overdose deaths, as nearly 45,000 individuals died from an opioid overdose in the United States in 2018 and over 4000 opioid-related deaths occurred in Canada in the same year [2]. The current system of addiction care has to deal with several challenges in providing care, especially for the most vulnerable populations. One such major challenge of the system of care is the availability of resources, which is limited in respect to incidence, prevalence, and distribution of substance use conditions [3]. Many also experience barriers in accessing treatment due to discrimination and stigma, particularly those in correctional settings, ethnic minorities, immigrants, and refugees [1]. In rural areas, people do not have sufficient access to specialist care. In the Yukon, for example, assuming a psychiatrist-to-population ratio of 1:10,000 [4], there should be between 4 and 5 psychiatrists. Currently, there are only 2 full-time general psychiatrists in the territory (ie, 1.6 per 100,000 population). Nunavut and the Northwestern Territories do not have any practicing psychiatrists [5]. Given the pressing need for effective measures to tackle this public health crisis, all stakeholders are increasingly committed to developing and implementing innovative solutions.

eHealth is a major field that has attracted growing attention for its versatility and accessibility. These interventions are delivered via electronic devices, such as computers, tablets, smartphones, and other handheld devices; delivery modes include websites, email, mobile apps, text messages, and telephone calls [6]. In this review, we distinguish between web-based interventions, intended for use through an internet browser on a computer; computer-based interventions (ie, software that is run offline on a computer, for example, from a CD or DVD); and apps (ie, native iOS and Android apps). These eHealth interventions have been used to target a range of substance use problems and have demonstrated effectiveness in previous meta-analyses [7-15]. However, the development of eHealth interventions for illicit drug use is still at a more formative stage.

Potential advantages of eHealth interventions are their wide reach, particularly among subpopulations such as young people and people residing in rural areas; higher likelihood of disclosing information due to anonymity, especially given the sensitive nature of substance use; lower maintenance cost, especially for automated self-help interventions; and easy transferability of the interventions to other languages and settings [16]. Additionally, fidelity is assured if educational material is delivered via the internet, as the material is delivered in its entirety and not adapted to a teacher's or counselor's style of delivery. This is not an uncommon problem and may cause removal of essential components of a program. Programs may also be expanded to other settings and still delivered in a consistent manner without the need for additional resources. Despite these disadvantages, many challenges with eHealth interventions can be avoided or mitigated by identifying the optimal eHealth intervention for a certain setting and context.

Given the increasing use of eHealth interventions to tackle the overdose crisis, decision makers need to make informed decisions about future investments in eHealth for substance use. Decisions must be based on several layers of complexity. First, eHealth interventions have substantial heterogeneity in the nature of the intervention (eg, information, psychoeducation, cognitive behavioral therapy [CBT], peer based, etc), mode of delivery (eg, website vs app), level of therapist involvement (eg, automated vs therapist assisted), and price (eg, free vs paid). Second, contextual factors, such as targeted subpopulation (eg, at-risk youth, inmates, and individuals with chronic substance use disorders [SUDs]) and implementation strategy (eg, part of school curriculum, part of a rehabilitation program, and freely available online), play an important role in the outcomes of eHealth interventions. Finally, the time frame for evaluating the desired outcomes should also be determined.

Traditional evaluations and systematic reviews tend to predominantly focus on whether the programs "worked" (eg, reduced substance use), often without an understanding of the complexity of the intervention in terms of for whom they may or may not have been effective, under what circumstances, and why. The realist review methodology used in this study can answer questions with this level of complexity. In this case, the basic evaluative question "What works?" changes to "What is it about this program that works, for whom, and under what circumstances?" [17]. Mechanisms matter in this approach because they explain the process that leads to outcomes, and the context is important because it changes the processes by which an intervention produces an outcome. Therefore, both context and mechanism must be systematically examined along with the intervention's strategy and outcome.

Realist reviews tend to start their synthesis with a wider scope and focus it during the review's iterations. In this review, all eHealth interventions targeting substance use were explored in the first iteration, and eHealth interventions based on CBT (ie, electronic CBT [eCBT]) were chosen as the scope of the realist review in the subsequent iterations. This was due to CBT being one of the most common evidence-based interventions that can be effectively delivered in an eHealth setting, among various approaches for substance use. In order to unpack the underlying mechanisms of how specific intervention strategies work in specific contexts, an important first step in a realist review is to narrow down, clarify, and refine the scope and, thus, the review question [17]. The aim of this realist review was to answer the following question: "How do different eCBT interventions for substance use interact with different contexts to produce certain outcomes?"

Methods

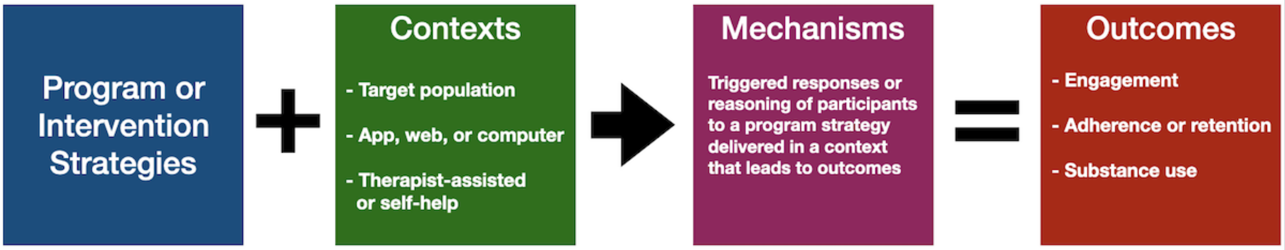
Realist Methodology

The methodology for this review has been adapted from Pawson et al [17], with an additional round of scoping to refine and clarify the research question in order to be able to conduct a feasible realist synthesis. Our research question stems from the following fundamental realist question: "What works for whom and why?" Specifically, "In an environment in which certain *contextual factors (C)* are at play, program activities or *strategies (S)* are implemented. These strategies, in combination with the contextual factors, trigger certain internal processes or changes in the participants' way of thinking, which is the *mechanism (M)* that, in turn, triggers the desired *outcome (O)*" [18]. Identifying program strategies and the mechanisms they trigger, alongside the context in which they are implemented, will allow us to generate knowledge that is useful for the implementation of programs and research that regulate outcomes in specific contexts. One intervention has multiple program strategies and each can trigger one or more mechanisms in different contexts that cause desirable or undesirable outcomes [17].

In this review, we also made use of the reconceptualization proposed by Dalkin et al [19], who argue that realist syntheses have difficulty deciding which aspects of an intervention contribute contextually or mechanistically to a Context-Mechanism-Outcome configuration (CMOc). It is important for both aspects to come together to be able to fully explain a program from a realist perspective.

By adding program strategy (ie, mechanism resource) as a variable between context and mechanism response (see Figure 1), we can be more accurate in how we construct our CMOcs and ensure that the mechanism is described more completely in terms of which resource triggers which response [19].

Figure 1. Overview of results according to the Context-Mechanism-Outcome formula proposed by Dalkin et al [19].



Search

We conducted a search on two electronic medical databases—MEDLINE and Embase—on June 26, 2018. The results were limited by published year (ie, 2009 to present) and by language (ie, English). The three main reasons for initially limiting the scope of the literature search in this study were as follows: (1) the tendency of realist reviews to develop a creeping nature consuming excessive amounts of time and resources, (2) the time-sensitive nature of the review given the relevance of its results to the opioid crisis, and (3) limited available resources. Moreover, we restricted the literature search to the past 10 years to ensure inclusion of current and relevant eHealth programs. The search strategy consisted of three baskets of key terms: substance use– and addiction-related keywords, eHealth- and web-related keywords, and the three keywords prevention, intervention, and treatment. Keywords under each group were combined with the "OR" operator, while the groups were merged with the "AND" operator. The three baskets helped in limiting results to find papers that focused on interventions for substance use using web-, computer-, or mobile-based components. The search returned 3184 records (see Multimedia Appendix 1). A gray literature search using Google and the search string "web-based substance intervention" was limited to the first 500 results and was completed on July 3, 2018.

Screening

Two reviewers independently screened the titles and abstracts of all database results. Disagreements were discussed between the reviewers and if not resolved, a third party in the research team was consulted. Inclusion criteria were as follows: (1) the paper examined web- or computer-based interventions, (2) the intervention targeted substance use (ie, alcohol, tobacco, cannabis, opioids, and stimulants), (3) the intervention tool was not being used for an electronic medical record or an electronic health record, (4) the intervention was not delivered via a wearable device (eg, watch or pedometer), and (5) if the intervention had telephone calls or text messaging as part of the intervention, it was also accompanied by a web-based component. Interventions that targeted only mental health conditions without a substance use component were not included. After removal of duplicates and screening, 186 records were included: 161 from MEDLINE and Embase and 25 from the Google search. References of key papers were tracked, and experts were consulted to identify additional relevant articles that were missed by the search. An additional 55 articles were included through reference tracking and expert consultation, resulting in a total of 241 articles being reviewed in full.

Quality Appraisal and Data Extraction

Data extraction began with the use of an appraisal form to ensure the fit of the papers for this review by examining rigor and relevance (see Multimedia Appendix 2). The form was adapted from the National Collaborating Centre for Methods and Tools appraisal form, in addition to consultation with the research team by pretesting the form to ensure its utility. Initially, the team also appraised two articles together and discussed the results as a team to ensure a consistent approach for this process. In realist reviews, the appraisal process is not based on the standardized evaluation of methodology that is typically used in systematic reviews (ie, randomized controlled trial studies possess the highest methodological standard) [17]. Rather, the appraisal process is based on the reviewer's judgement of the fitness of an article for that specific synthesis [17]. The assessment of methodological soundness and trustworthiness of the paper is based on the appropriateness of the methodology with what was stated as the goal or objective of the study. This rigor was assessed using the following questions: (1) Did the authors have clearly defined research goals? (2) Were the research goals adequately addressed in the results and discussion sections of the article? (3) Was the methodology used in the study appropriate with respect to the research goals? and (4) Overall, did the study have reliability and validity? Following these criteria, a study may be excluded for reasons such as the sample size being too small or too homogeneous (eg, only female). Relevance was ensured by considering whether the paper had a direct relevance to our research project by contributing to forming the program theory. During this process, 18 articles were excluded because they did not meet rigor and/or relevance requirements. Furthermore, after excluding study protocols and those articles that were not accessible, 208 articles met the initial review inclusion criteria and were used for data extraction.

Data extraction was completed with a data extraction template that collected information on intervention details, study setting and context, results, and mechanisms. During this process, it became clear that the scope of the synthesis was too broad, given the limited time and resources available. To address that, articles that were of the *review* type (ie, systematic reviews), protocols, and cost-effectiveness studies were excluded from the synthesis.

After preliminary data extraction, we further narrowed down the scope to only focus on CBT-based programs. First, all studies that looked at eHealth interventions that were based on a CBT theoretical framework were identified. Second, studies were excluded if (1) they solely addressed alcohol and/or tobacco use with no other substance use, (2) they were review papers, and (3) they did not identify or discuss any mechanisms of change. Alcohol and tobacco use eHealth interventions have a large body of literature devoted to them and they tend to target a wide array of populations, which may result in very heterogeneous contexts compared to other substance use interventions. This study sought to narrow down its focus to illicit substances, therefore excluding nicotine and alcohol. It is important to note that recreational cannabis use was still illegal in Canada at the time of the review and is still illegal in most countries around the world, including in the United States under federal law; therefore, these studies were included in this review. A total of 54 articles met the eligibility criteria after refining the scope and were included in the final synthesis.

Context-Mechanism-Outcome Configurations

In order to study how different strategies and contexts produce particular outcomes, each reviewer extracted candidate CMOcs based on qualitative discussions found in the identified literature that described how an intervention or parts of an intervention may or may not work. The context and outcomes of the paper were summarized in order to create full CMOcs that included a program strategy, context, mechanism, and outcome. After identifying all candidate CMOcs, along with their program strategies, the quality of each CMOc was appraised using the following criteria: (1) How rigorous and specific was the CMOc? (2) Was it possible to make a clear distinction between program strategies (ie, mechanism resource) and mechanism response? and (3) How relevant was the CMOc in improving the understanding of, and providing strategies to improve, eCBT interventions for substance use? Reviewers analyzed studies for themes in the candidate CMOcs using keywords identified in the mechanisms, outcomes, and program strategies. Four reviewers rated and discussed the quality of the candidate CMOcs (ie, rigor and relevance) and, in consultation with the principal investigator, candidate CMOcs were reduced to a final set.

Results

A total of 54 papers from 24 unique programs that are grounded in CBT theory were reviewed. Eight CMOcs from five unique programs met the inclusion criteria and were included in the final synthesis.

Table 1 summarizes the eight CMOcs that were included from the reviewed literature regarding eCBT interventions [20-31] . The study setting, which includes how and where the study was carried out, is an important part of the intervention context. For example, the Computer-Based Training for Cognitive Behavioral Therapy (CBT4CBT) studies were delivered as a blended model combining CBT4CBT with treatment as usual, such as methadone maintenance treatment for opioid use disorder, as well as having a research associate available at all times to answer questions related to the program. The Self-Help for Alcohol and Other Drug Use and Depression (SHADE) and CBT for insomnia (CBT-I) interventions, on the other hand, were delivered to participants with the assistance of therapists or clinicians. Out of the five programs, two were delivered in a web-based format, two were computer-based interventions (eg, using CDs or DVDs in an offline context), and one was delivered as an app. Most interventions were targeting individuals with SUDs.

Table 1. Summary of components of the eight Context-Mechanism-Outcome configurations (CMOcs).

CMOc No.	Context of intervention	Program strategy	Mechanism	Outcome	Program
1	Individuals with substance use disorder (SUD) No human involvement; self-help electronic cognitive behavioral therapy (eCBT)	The program focused on strengthening coping skills	Improvement in coping skills, strengthening of executive cognitive control, and reduction of attentional bias toward drug-related cues	Reduced cue-induced craving, resulting from a reduction of attentional bias Retention in treatment Drug use and abstinence	CBT4CBT[a] [20-23]
2	Individuals with SUD and possible cognitive impairment No human involvement; self-help eCBT	The program material was prepared for higher literacy levels and normal levels of cognitive functioning	Negative effect on self-efficacy and commitment to abstinence, specifically in people with cognitive impairment	Reduced retention in treatment Higher dropout rates	CBT4CBT [24,25]
3	Individuals with SUD accessing Breaking free Online (BFO) drug and alcohol treatment services across the United Kingdom No human involvement; self-help eCBT	The program provided tailored feedback on the level of impairment on each module	Recognition of problem areas	Increased use of modules that were highlighted red (ie, highest level of impairment)	BFO [26]
4	Individuals with SUD accessing BFO drug and alcohol treatment services across the United Kingdom No human involvement; self-help eCBT	Completion of a higher *dose* of modules was facilitated by the program	*Dosage effect* (ie, more intervention tasks completed will lead to improved outcomes)	Increased improvement in functioning	BFO [26]
5	Individuals with SUD accessing BFO drug and alcohol treatment services across the United Kingdom No human involvement; self-help eCBT	Participants were given cognitive restructuring training	Recognition of dysfunctional beliefs	Reduced severity of alcohol dependence	BFO [26]
6	Individuals with current comorbid depression and problematic alcohol and cannabis use Therapist clinician–assisted eCBT	The program was designed as a self-help intervention	Feelings of empowerment and possible enhancement in problem-solving skills	Increased client initiative and acceptability	SHADE[b] [27,28]
7	Veterans with cannabis use disorder and current sleep problems Therapist clinician–assisted eCBT	Participants were reminded to track behavior	Feelings of accountability	Increased adherence to tracking behavior	CBT-I[c] Coach [29,30]
8	Veterans No human involvement; self-help eCBT	Users participated in focus groups and individual feedback sessions to inform development of the program	Relatability and relevance of content to target population (ie, veterans)	Improved likability, ease of use, and relevance of the program Increased engagement	CHMF[d] [31]

[a] CBT4CBT: Computer-Based Training for Cognitive Behavioral Therapy.

[b] SHADE: Self-Help for Alcohol and Other Drug Use and Depression.

[c] CBT-I: cognitive behavioral therapy for insomnia.

[d] CHMF: Coming Home and Moving Forward.

CBT4CBT is a computer-based program to help people stop or reduce the use of substances, including alcohol. The two studies, from four papers, included in this synthesis had 77 and 101 participants with mean ages of 40.6 and 43.1 years, respectively. A total of 40% and 60% of participants were female, respectively. We found that CBT4CBT, provided as part of a blended model—treatment as usual (eg, methadone treatment + CBT4CBT)—of substance use treatment in the clinic for people with SUD, can improve retention in treatment and abstinence by strengthening participants' coping skills and executive cognitive control, as well as reducing attentional bias (ie, CMOc 1). The authors associate a reduction in attentional bias with a reduction in cue-induced craving [20-23].

Additionally, for CBT4CBT, two additional studies were included. Participants in the two studies had a mean age of 42.3 and 38.3 years, respectively. Sample sizes were 52 and 120, of which 40% and 33.3% were female, respectively [23,24]. We found that cognitive impairment had a negative influence on retention in treatment through affecting participant change mechanisms, such as self-efficacy and commitment to abstinence (ie, CMOc 2). Preparing material for low literacy levels can possibly address this issue to some extent [32].

Breaking Free Online (BFO) is a web-based relapse-prevention program that targets individuals with SUD. One major study with a large sample (N=2311) was included in our synthesis and resulted in three CMOcs. The mean age of the participants in this study was 42.2 years (range 15-76) and 45% of the participants were female. We found that providing tailored symptom feedback on biopsychosocial impairment (see Lifestyle Balance Model in Davies et al [33]) that highlights areas the user needs to work on will increase usage of those modules by the user, because users are able to recognize the importance of working on those areas with significant levels of impairment (ie, CMOc 3) [26].

In the BFO study, we also found that the greater number of modules completed, the greater the improvement reported in functioning, due to the so-called *dosage effect* (ie, CMOc 4). The dosage effect indicates that an increase in the number of tasks or intervention strategies completed will lead to improved outcomes; in this case, a greater improvement in biopsychosocial functioning [26].

Finally, in the BFO study, we found that in an eCBT intervention, cognitive restructuring can help lower the severity of alcohol dependence because it targets “dysfunctional beliefs that underpin and maintain unhelpful and unhealthful behaviours” (ie, CMOc 5) [26].

The SHADE intervention is a computer-delivered motivational treatment for individuals with co-occurring SUD and depression. The mean age of the participants in the SHADE study was 35.37 years (range 18-61), and 54% of the participants were female in a total sample of 97. In the therapist clinician–delivered CBT group, 19 out of 35 (54%) participants completed all sessions. In the computer-delivered group, 15 out of 32 (47%) participants completed all sessions. A total of 30 participants were not assigned to any treatment group. The self-help nature of the program, which allows users to take charge of their own intervention, increased client initiative and, thus, treatment acceptability through a sense of empowerment and enhanced problem-solving skills (ie, CMOc 6). Client initiative is one of the subcategories of treatment acceptability and describes the extent to which a client feels that they are directing therapy. The authors suggest that higher levels of client initiative are associated with changes in alcohol use. It has also been suggested that the self-help nature of a treatment leads to more sustainable outcomes in the longer term [27,28].

CBT-I Coach is a mobile app designed to be an adjunct to in-person CBT-I. Two studies of CBT-I Coach were included. The first study had 4 male veterans with a mean age of 47 years (SD 16.31, range 27-65). The veterans diagnosed with DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) cannabis use disorder and sleep problems were randomized to receive a 2-week intervention with (1) the CBT-I Coach mobile app (n=2) or (2) a placebo, control mood-tracking app (n=2). In the second study, the participants’ mean age was 48.50 years (SD 14.93). The total sample size was 18 and 39% of the participants were female. The CBT-I Coach studies suggest that including regular reminders (eg, daily notifications) can help increase adherence to the tracked behavior, because individuals feel like they are being held accountable to the program (ie, CMOc 7) [29,30].

Coming Home and Moving Forward (CHMF) is a web-based self-help program for recent combat veterans with posttraumatic stress disorder and substance use. Mean age and range were not reported, but veterans were either part of focus groups (n=18) or individual feedback sessions (n=34). Each group had 4 female participants. The veterans that participated in this study found a web-based intervention based on CBT principles to be likable, easy to use, and relevant to their experiences, because focus group feedback helped to make the intervention more contextually anchored in their experiences (CMOc 8) [31].

Discussion

Principal Findings

In this review, we sought to answer the following question based on the original realist review question: “How do different eCBT interventions for substance use interact with different contexts to produce certain outcomes?” The review revealed some strategies that could be used in some contexts in order to improve substance use outcomes. The results of the review identified five strategies that may be considered when developing or implementing an eCBT intervention targeting substance use.

The first strategy is to consider addressing cognitive functioning and/or impairment. There is a body of evidence on the potential impact of cognitive functioning on treatment response and substance use outcomes in web- or computer-based CBT interventions [34-37]. Some authors suggest that more than 50% of people entering substance use treatment have some level of impairment [37-39]. Assessing neuropsychological functioning and tailoring content to the appropriate level of cognitive functioning is, therefore, important in order to improve understanding of the material. Simplifying content or targeting impairment directly via cognitive training are two other strategies [40,41]. The program strategy identified in CMOc 3 may be applied to this kind of impairment as well.

The second strategy is to tailor content to the user’s needs. Tailored content may help improve engagement and other treatment outcomes [42]. One strategy identified in a program that addressed cognitive functioning (ie, CMOc 3) assessed levels of impairment in the aspects of the Lifestyle Balance Model [33] and provided participants with tailored feedback, highlighting the most important modules to work on. This resulted in highlighted modules being worked on more, and those modules resulted in greater improved treatment outcomes in comparison to modules that were worked on less [26].

Third, cognitive restructuring has been effectively used in dealing with negative thoughts [43]. There is evidence that the delivery of cognitive restructuring material can be effective in a digital context. In a study conducted on BFO [26], completion of the *negative thoughts*, cognitive restructuring, Lifestyle Balance Model intervention strategies was associated with improved outcomes across all measures, including severity of alcohol dependence, depression, anxiety, and quality of life.

Fourth, our review revealed that despite increasing focus on including users’ experiences and perspectives in the development of new solutions, existing solutions were more often adaptations and digitalization of conventional treatment approaches. This does not leave much room for involvement of users in forming the digital solution. One program targeting veterans used focus groups to make the intervention more tailored to the veterans' needs. As a result, veterans described the program as likable, easy to use, and relevant to their experiences [31]. This shows the importance of including the target population in the early stages of the intervention design.

Finally, engagement is a key ingredient that needs attention in order to increase usage, retention, and adherence to an eHealth intervention [42,44]. This review identified multiple strategies that can be used to improve engagement of users with the program. The *dosage effect* explains that an increase in usage of modules will lead to greater improvement in desired outcomes [26]. Facilitating the completion of intervention strategies and modules should be a key focus of eCBT interventions. Additionally, if utilized well, the self-help nature of some interventions may be an advantage, as it may increase client initiative and, thus, acceptability of the program (ie, CMOc 6). Making sure additional criteria for acceptability are met, such as ease of use and an intuitive system design (eg, by applying cocreation principles in the design and development of interventions), can further help to improve engagement and other treatment outcomes. As described in CMOc 7, timely reminders can also play an important role in improving engagement with programs.

Limitations

The majority of publications in this field reported on randomized controlled trials. An issue with conducting a realist review on these types of publications is that their primary focus is on quantitative results without any attempt to interpret the mechanisms. Thus, any realist review must generally focus on screening of the discussion section of publications to identify author opinions or any qualitative information that may shine a light on the mechanisms of how certain interventions work. Very few of the studies published in the literature on eHealth CBT interventions for substance use provided enough information to investigate the theory and the mechanisms that drove the intervention outcomes in different contexts. As such, most of the information was inferred from an appraisal of the authors' comments, and subjectivity of these inferences must be acknowledged. Only one study statistically analyzed mechanisms of action [26].

Another limitation is the lack of reporting of more proximal outcomes in the studies; for example, retention or adherence outcomes, changes in perception, willingness to change, and self-efficacy. Although long-term goals, such as reducing substance use, are crucial for understanding effectiveness of interventions, proximal outcomes may be of help when trying to understand the mechanisms of action behind the effects.

Conclusions

To our knowledge, this is the first published realist review that aims to identify what works for whom in eCBT interventions that target SUD. The review identified eight CMOcs from five programs, and we discussed five strategies that we have identified through synthesis of the findings from the CMOcs: addressing cognitive functioning, tailoring content to user needs, addressing negative thoughts, cocreation, and addressing engagement. These strategies, among others, may be considered when developing or making funding decisions regarding eCBT programs for SUD. eCBT interventions are an increasingly popular tool for program developers for addressing substance use problems, yet additional research is needed to identify what factors lead to desired outcomes and to explain the mechanisms of change. Thus, future studies should focus on clearly defining the components of an intervention, recording these mechanisms, and defining their relationship to outcomes and context, in order to better understand what components of a program work and why.

Acknowledgments

We thank Dr Justin Jagosh, the University of Liverpool, for his guidance on the realist review methodology; Vanessa Kitchin, the University of British Columbia, for her guidance with the literature search; and Dr Amy Porath and Dr Sheena Taha, the Canadian Centre on Substance Use and Addiction, for supporting our work as knowledge users. This work was supported by the Canadian Institutes of Health Research Operating Grant: Opioid Crisis Knowledge Synthesis (funding reference No. 348103 - 398000). MN was supported by the Frederick Banting and Charles Best Canada Graduate Scholarships Award (funding reference No. 157934).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Initial literature review search strategy.
[DOCX File , 19 KB - jmir_v23i1e20557_app1.docx]

Multimedia Appendix 2
Appraisal form.
[DOC File , 36 KB - jmir_v23i1e20557_app2.doc]

References

1. United Nations Office on Drugs and Crime. World Drug Report 2020. New York, NY: United Nations; 2020.

2. The drug overdose epidemic: Behind the numbers. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/drugoverdose/data/index.html [accessed 2021-04-01]
3. Malla A, Shah J, Iyer S, Boksa P, Joober R, Andersson N, et al. Youth mental health should be a top priority for health care in Canada. Can J Psychiatry 2018 Dec;63(4):216-222 [FREE Full text] [doi: 10.1177/0706743718758968] [Medline: 29528719]
4. Burvill P. Looking beyond the 1:10,000 ratio of psychiatrists to population. Aust N Z J Psychiatry 1992 Jun;26(2):265-269. [Medline: 1642618]
5. Psychiatry Profile. Ottawa, ON: Canadian Medical Association; 2019 Dec. URL: https://www.cma.ca/sites/default/files/2019-01/psychiatry-e.pdf [accessed 2021-01-04]
6. Muellmann S, Forberger S, Möllers T, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: A systematic review protocol. Syst Rev 2016 Mar 16;5:47 [FREE Full text] [doi: 10.1186/s13643-016-0223-7] [Medline: 26984189]
7. Boumparis N, Karyotaki E, Schaub MP, Cuijpers P, Riper H. Internet interventions for adult illicit substance users: A meta-analysis. Addiction 2017 Sep;112(9):1521-1532 [FREE Full text] [doi: 10.1111/add.13819] [Medline: 28295758]
8. Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. Am J Psychiatry 2008 Feb;165(2):179-187. [doi: 10.1176/appi.ajp.2007.06111851] [Medline: 18198270]
9. Powers MB, Vedel E, Emmelkamp PMG. Behavioral couples therapy (BCT) for alcohol and drug use disorders: A meta-analysis. Clin Psychol Rev 2008 Jul;28(6):952-962. [doi: 10.1016/j.cpr.2008.02.002] [Medline: 18374464]
10. Roozen HG, Boulogne JJ, van Tulder MW, van den Brink W, De Jong CAJ, Kerkhof AJFM. A systematic review of the effectiveness of the community reinforcement approach in alcohol, cocaine and opioid addiction. Drug Alcohol Depend 2004 Apr 09;74(1):1-13. [doi: 10.1016/j.drugalcdep.2003.12.006] [Medline: 15072802]
11. Smedslund G, Berg RC, Hammerstrøm KT, Steiro A, Leiknes KA, Dahl HM, et al. Motivational interviewing for substance abuse. Cochrane Database Syst Rev 2011(5):CD008063. [doi: 10.1002/14651858.CD008063.pub2] [Medline: 21563163]
12. Magill M, Ray LA. Cognitive-behavioral treatment with adult alcohol and illicit drug users: A meta-analysis of randomized controlled trials. J Stud Alcohol Drugs 2009 Jul;70(4):516-527 [FREE Full text] [Medline: 19515291]
13. Lussier JP, Heil SH, Mongeon JA, Badger GJ, Higgins ST. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. Addiction 2006 Feb;101(2):192-203. [doi: 10.1111/j.1360-0443.2006.01311.x] [Medline: 16445548]
14. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: A meta-analysis. PLoS One 2014;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
15. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
16. Tossmann HP, Leuschner F. Internet-Based Drug Treatment Interventions: Best Practice and Applications in EU Member States. Lisbon, Portugal: European Monitoring Centre for Drugs and Drug Addiction (EMCDDA); 2009 Jul. URL: https://www.emcdda.europa.eu/system/files/publications/544/EMCDDA-Insight10_155820.pdf [accessed 2021-01-10]
17. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--A new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10 Suppl 1:21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]
18. Velonis A, Mahabir D, Maddox R, O'Campo P. Still looking for mechanisms: A realist review of batterer intervention programs. Trauma Violence Abuse 2020 Oct;21(4):741-753. [doi: 10.1177/1524838018791285] [Medline: 30122126]
19. Dalkin SM, Greenhalgh J, Jones D, Cunningham B, Lhussier M. What's in a mechanism? Development of a key concept in realist evaluation. Implement Sci 2015 Apr 16;10:49 [FREE Full text] [doi: 10.1186/s13012-015-0237-x] [Medline: 25885787]
20. Sugarman DE, Nich C, Carroll KM. Coping strategy use following computerized cognitive-behavioral therapy for substance use disorders. Psychol Addict Behav 2010 Dec;24(4):689-695 [FREE Full text] [doi: 10.1037/a0021584] [Medline: 21198228]
21. Kiluk BD, DeVito EE, Buck MB, Hunkele K, Nich C, Carroll KM. Effect of computerized cognitive behavioral therapy on acquisition of coping skills among cocaine-dependent individuals enrolled in methadone maintenance. J Subst Abuse Treat 2017 Nov;82:87-92 [FREE Full text] [doi: 10.1016/j.jsat.2017.09.011] [Medline: 29021121]
22. Kiluk B, Nich C, Babuscio T, Carroll K. Quality versus quantity: Acquisition of coping skills following computerized cognitive-behavioral therapy for substance use disorders. Addiction 2010 Dec;105(12):2120-2127 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.03076.x] [Medline: 20854334]
23. DeVito EE, Kiluk BD, Nich C, Mouratidis M, Carroll KM. Drug Stroop: Mechanisms of response to computerized cognitive behavioral therapy for cocaine dependence in a randomized clinical trial. Drug Alcohol Depend 2018 Feb 01;183:162-168 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.10.022] [Medline: 29258028]
24. Kiluk BD, Nich C, Carroll KM. Relationship of cognitive function and the acquisition of coping skills in computer assisted treatment for substance use disorders. Drug Alcohol Depend 2011 Apr 01;114(2-3):169-176 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.09.019] [Medline: 21050679]

25. Carroll KM, Nich C, DeVito EE, Shi JM, Sofuoglu M. Galantamine and computerized cognitive behavioral therapy for cocaine dependence: A randomized clinical trial. J Clin Psychiatry 2018;79(1):1-18 [FREE Full text] [doi: 10.4088/JCP.17m11669] [Medline: 29286595]
26. Elison S, Jones A, Ward J, Davies G, Dugdale S. Examining effectiveness of tailorable computer-assisted therapy programmes for substance misuse: Programme usage and clinical outcomes data from Breaking Free Online. Addict Behav 2017 Nov;74:140-147. [doi: 10.1016/j.addbeh.2017.05.028] [Medline: 28645092]
27. Kay-Lambkin F, Baker A, Lewin T, Carr V. Acceptability of a clinician-assisted computerized psychological intervention for comorbid mental health and substance use problems: Treatment adherence data from a randomized controlled trial. J Med Internet Res 2011 Jan 27;13(1):e11 [FREE Full text] [doi: 10.2196/jmir.1522] [Medline: 21273184]
28. Kay-Lambkin FJ, Baker A, Lewin T, Carr V. Computer-based psychological treatment for comorbid depression and problematic alcohol and/or cannabis use: A randomized controlled trial of clinical efficacy. Addiction 2009 Mar;104(3):378-388. [doi: 10.1111/j.1360-0443.2008.02444.x] [Medline: 19207345]
29. Babson K, Ramo D, Baldini L, Vandrey R, Bonn-Miller M. Mobile app-delivered cognitive behavioral therapy for insomnia: Feasibility and initial efficacy among veterans with cannabis use disorders. JMIR Res Protoc 2015 Jul 17;4(3):e87 [FREE Full text] [doi: 10.2196/resprot.3852] [Medline: 26187404]
30. Koffel E, Kuhn E, Petsoulis N, Erbes C, Anders S, Hoffman J, et al. A randomized controlled pilot study of CBT-I Coach: Feasibility, acceptability, and potential impact of a mobile phone application for patients in cognitive behavioral therapy for insomnia. Health Informatics J 2018 Mar;24(1):3-13 [FREE Full text] [doi: 10.1177/1460458216656472] [Medline: 27354394]
31. Possemato K, Acosta MC, Fuentes J, Lantinga LJ, Marsch LA, Maisto SA, et al. A web-based self-management program for recent combat veterans with PTSD and substance misuse: Program development and veteran feedback. Cogn Behav Pract 2015 Aug 01;22(3):345-358 [FREE Full text] [doi: 10.1016/j.cbpra.2014.03.005] [Medline: 26120269]
32. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Nuro KF, et al. Computer-assisted delivery of cognitive-behavioral therapy for addiction: A randomized trial of CBT4CBT. Am J Psychiatry 2008 Jul;165(7):881-888 [FREE Full text] [doi: 10.1176/appi.ajp.2008.07111835] [Medline: 18450927]
33. Davies G, Elison S, Ward J, Laudet A. The role of lifestyle in perpetuating substance use disorder: The Lifestyle Balance Model. Subst Abuse Treat Prev Policy 2015 Jan 17;10:2 [FREE Full text] [doi: 10.1186/1747-597X-10-2] [Medline: 25595205]
34. Aharonovich E, Nunes E, Hasin D. Cognitive impairment, retention and abstinence among cocaine abusers in cognitive-behavioral treatment. Drug Alcohol Depend 2003 Aug 20;71(2):207-211 [FREE Full text] [doi: 10.1016/s0376-8716(03)00092-9] [Medline: 12927659]
35. Aharonovich E, Hasin DS, Brooks AC, Liu X, Bisaga A, Nunes EV. Cognitive deficits predict low treatment retention in cocaine dependent patients. Drug Alcohol Depend 2006 Feb 28;81(3):313-322. [doi: 10.1016/j.drugalcdep.2005.08.003] [Medline: 16171953]
36. Bates ME, Pawlak AP, Tonigan JS, Buckman JF. Cognitive impairment influences drinking outcome by altering therapeutic mechanisms of change. Psychol Addict Behav 2006 Sep;20(3):241-253 [FREE Full text] [doi: 10.1037/0893-164X.20.3.241] [Medline: 16938062]
37. Morgenstern J, Bates ME. Effects of executive function impairment on change processes and substance use outcomes in 12-step treatment. J Stud Alcohol 1999 Nov;60(6):846-855. [doi: 10.15288/jsa.1999.60.846] [Medline: 10606498]
38. Meek PS, Clark HW, Solana VL. Neurocognitive impairment: The unrecognized component of dual diagnosis in substance abuse treatment. J Psychoactive Drugs 1989;21(2):153-160. [doi: 10.1080/02791072.1989.10472155] [Medline: 2760756]
39. O'Malley S, Adamse M, Heaton RK, Gawin FH. Neuropsychological impairment in chronic cocaine abusers. Am J Drug Alcohol Abuse 1992;18(2):131-144. [doi: 10.3109/00952999208992826] [Medline: 1562011]
40. Aharonovich E, Hasin DS, Nunes EV, Stohl M, Cannizzaro D, Sarvet A, et al. Modified cognitive behavioral therapy (M-CBT) for cocaine dependence: Development of treatment for cognitively impaired users and results from a Stage 1 trial. Psychol Addict Behav 2018 Nov;32(7):800-811 [FREE Full text] [doi: 10.1037/adb0000398] [Medline: 30346186]
41. Sofuoglu M, DeVito EE, Waters AJ, Carroll KM. Cognitive enhancement as a treatment for drug addictions. Neuropharmacology 2013 Jan;64:452-463 [FREE Full text] [doi: 10.1016/j.neuropharm.2012.06.021] [Medline: 22735770]
42. Couper M, Alexander G, Zhang N, Little R, Maddy N, Nowak M, et al. Engagement and retention: Measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
43. Larsson A, Hooper N, Osborne LA, Bennett P, McHugh L. Using brief cognitive restructuring and cognitive defusion techniques to cope with negative thoughts. Behav Modif 2016 May;40(3):452-482. [doi: 10.1177/0145445515621488] [Medline: 26685210]
44. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]

Abbreviations

BFO: Breaking Free Online
CBT: cognitive behavioral therapy
CBT4CBT: Computer-Based Training for Cognitive Behavioral Therapy
CBT-I: cognitive behavioral therapy for insomnia
CHMF: Coming Home and Moving Forward
CMOc: Context-Mechanism-Outcome configuration
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
eCBT: electronic cognitive behavioral therapy
SHADE: Self-Help for Alcohol and Other Drug Use and Depression
SUD: substance use disorder

Edited by T Rashid Soron; submitted 21.05.20; peer-reviewed by M Tolou-Shams, J Folk; comments to author 30.06.20; revised version received 10.08.20; accepted 05.12.20; published 21.01.21.

Please cite as:
Shams F, Wong JSH, Nikoo M, Outadi A, Moazen-Zadeh E, Kamel MM, Song MJ, Jang KL, Krausz RM
Understanding eHealth Cognitive Behavioral Therapy Targeting Substance Use: Realist Review
J Med Internet Res 2021;23(1):e20557
URL: http://www.jmir.org/2021/1/e20557/
doi:10.2196/20557
PMID:33475520

Review

Possibilities, Problems, and Perspectives of Data Collection by Mobile Apps in Longitudinal Epidemiological Studies: Scoping Review

Florian Fischer[1,2,3], PhD; Sina Kleen[3], MSc

[1]Institut of Public Health, Charité - Universitätsmedizin Berlin, Berlin, Germany

[2]Institute of Gerontological Health Services and Nursing Research, Ravensburg-Weingarten University of Applied Sciences, Weingarten, Germany

[3]Department of Population Medicine and Health Services Research, School of Public Health, Bielefeld University, Bielefeld, Germany

Corresponding Author:
Florian Fischer, PhD
Institut of Public Health
Charité - Universitätsmedizin Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 751 501 9441
Email: florian.fischer1@charite.de

Abstract

Background: The broad availability of smartphones and the number of health apps in app stores have risen in recent years. Health apps have benefits for individuals (eg, the ability to monitor one's health) as well as for researchers (eg, the ability to collect data in population-based, clinical, and observational studies). Although the number of health apps on the global app market is huge and the associated potential seems to be great, app-based questionnaires for collecting patient-related data have not played an important role in epidemiological studies so far.

Objective: This study aims to provide an overview of studies that have collected patient data using an app-based approach, with a particular focus on longitudinal studies. This literature review describes the current extent to which smartphones have been used for collecting (patient) data for research purposes, and the potential benefits and challenges associated with this approach.

Methods: We conducted a scoping review of studies that used data collection via apps. PubMed was used to identify studies describing the use of smartphone app questionnaires for collecting data over time. Overall, 17 articles were included in the summary.

Results: Based on the results of this scoping review, there are only a few studies that integrate smartphone apps into data-collection approaches. Studies dealing with the collection of health-related data via smartphone apps have mainly been developed with regard to psychosomatic, neurodegenerative, respiratory, and cardiovascular diseases, as well as malign neoplasm. Among the identified studies, the duration of data collection ranged from 4 weeks to 12 months, and the participants' mean ages ranged from 7 to 69 years. Potential can be seen for real-time information transfer, fast data synchronization (which saves time and increases effectivity), and the possibility of tracking responses longitudinally. Furthermore, smartphone-based data-collection techniques might prevent biases, such as reminder bias or mistakes occurring during manual data transfers. In chronic diseases, real-time communication with physicians and early detection of symptoms enables rapid modifications in disease management.

Conclusions: The results indicate that using mobile technologies can help to overcome challenges linked with data collection in epidemiological research. However, further feasibility studies need to be conducted in the near future to test the applicability and acceptance of these mobile apps for epidemiological research in various subpopulations.

(J Med Internet Res 2021;23(1):e17691) doi:10.2196/17691

KEYWORDS

apps; questionnaire; survey; epidemiology; healthcare

Introduction

Worldwide, there were 2.39 billion smartphone users in 2017, and this number is predicted to exceed three billion by 2021 [1]. Undoubtedly, smartphones have become part of everyday life for many people around the world. Alongside the increase in smartphone use, the market for smartphone apps has risen massively in recent years [2]. In health care, it is anticipated that apps have the potential to decrease both spatial and administrative barriers [2]. However, removing barriers is only one potential outcome resulting from the use of these applications. The market for mHealth apps is growing rapidly. In 2015 alone, there were more than 100,000 new health apps published in relevant app stores, with approximately three billion downloads of apps related to health, fitness, and medicine [2]. In comparison to 2013, the number of downloaded apps has almost doubled [2]. Consumers use apps to track steps, heart rate, sleep patterns, and so forth. Smartphones also function as new tools for measuring the health of individuals. Moreover, smartphone apps may deliver data at the population level. For that reason, smartphones are expected to be beneficial to health care research and epidemiological studies. Over the last two decades, surveys have ceased to exclusively use traditional methods of data collection, such as paper or telephone-based questionnaires; questionnaires have instead been developed using electronic systems, such as internet-based surveys and personal digital assistants. Although there has been a huge rise in the use of smartphones, issues regarding the benefits and potential uses of app-based questionnaires still need to be addressed [3]. Various studies have developed smartphone apps for educational or communication purposes for medical school students and clinicians [4]. However, only a few questionnaires on smartphone apps have been used in clinical settings, such as in sleep disorder tracking and for the administration of psychiatric questionnaires [5-7]. Beyond epidemiology, public health, and health services research, there are approaches in the field of cognitive science that have also detected the potential for data collection via smartphone apps [8].

Although data collection using smartphone apps has not yet been comprehensively studied and is not very frequently used, the availability of and access to smartphones among the world's population raises the potential for large-scale surveys in (longitudinal) population-based studies. This scoping review focuses on studies investigating the acceptability, feasibility, and performance of mobile apps for data collection in longitudinal studies, irrespective of the studies' purposes (eg, measuring vital parameters, providing questionnaires) or the types of apps in question (eg, apps that are medical products prescribed by physicians or those that are lifestyle products selected by users). To guide the scoping review, we focused on the following 4 overarching research questions: 1) which types of studies (feasibility/pilot vs full-scale) have been conducted to date, 2) over what period of time has data been collected within these studies, 3) which specific target groups (eg, children, elderly, migrants) have been included in the studies, and 4) what potential benefits and challenges to app-based data collection are described in the studies?

Methods

We conducted a scoping review using PubMed. In the first step, we employed a broad search strategy to include all articles dealing with apps, smartphone devices, or mHealth in the context of data-collection activities. Hence, the following search algorithm was used:

> *(app[Title/Abstract] OR apps[Title/Abstract] OR smartphone*[Title/Abstract] OR mHealth[Title/Abstract] OR "mobile health"[Title/Abstract] OR mobile phone*[Title/Abstract]) AND (cohort*[Title/Abstract] OR survey*[Title/Abstract] OR questionnaire*[Title/Abstract])*

All articles published up to December 31, 2017 were considered by two independent reviewers. This led to 1922 matches after incorporating the predefined filters for studies conducted among humans and those written in the requisite publication language(s) (English or German). The following main exclusion criteria were defined a priori:

- SMS-based approaches (including SMS reminders and supportive text messages in interventional studies)
- Web-based approaches/questionnaires, which can also be conducted on smartphones
- Nomophobia (fear of being out of cellular phone contact)
- Distraction by smartphones while driving
- "App" as an abbreviation for unrelated issues (eg, amyloid precursor protein)
- Screen media time usage among children or adolescents
- Mobile phone access and usage in various populations
- Validation of paper-based compared with software/app-based questionnaires
- Articles dealing with Electronic Health Records
- mHealth interventions using apps only for treatment or educational purposes (eg, limited to disease treatments)

After screening all article titles and abstracts against the criteria listed above, 1808 articles were excluded. This led to 114 articles for which a review of the full text was performed. Due to the large number of articles still remaining, further exclusion criteria were defined and the particular reason for exclusion was documented. Using this procedure, 97 further articles were excluded for the following reasons:

- Cross-sectional study design [9-15]
- Qualitative study design [16]
- Study protocol [17]
- App without purpose of data collection [18-26]
- App only for self-monitoring or willingness to self-monitor disease or certain lifestyle measures (no data storage for researcher) [27-48]
- Mobile-type programs/programmed phones, personal monitors, or ecological momentary assessment protocols displaying on mobile phone screens (without the involvement of an app) [49-52]
- Advanced data-collection systems without the need for data entry by the patient/interviewee (eg, using Global

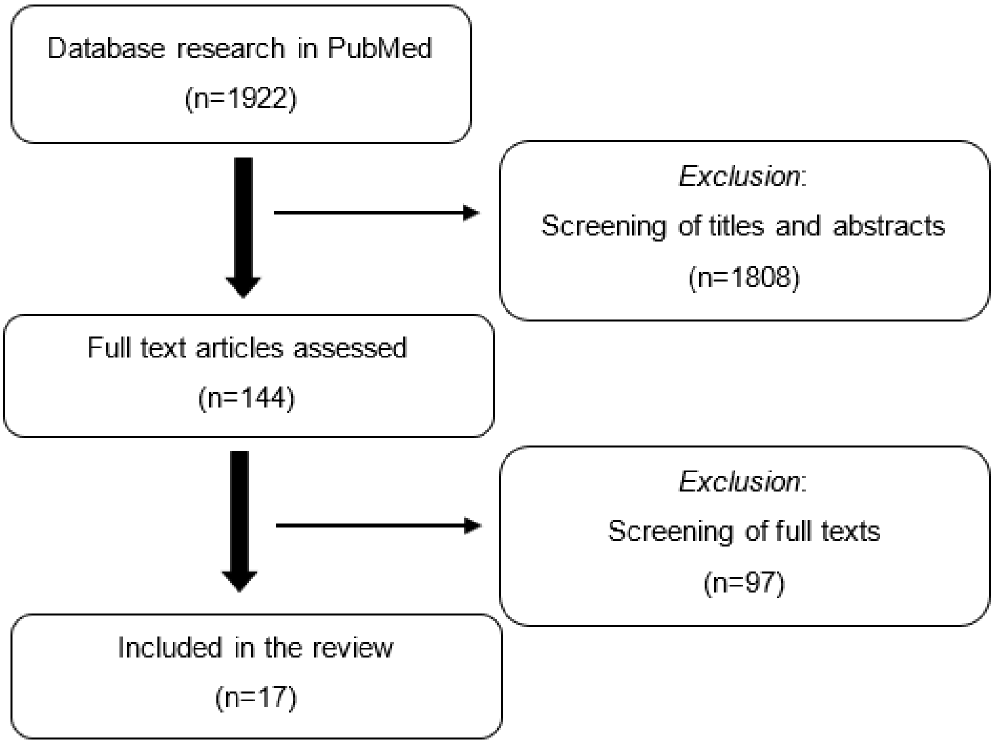
Positioning System location and phone usage data automatically) [53-55]
- App as assistive working tool for researchers or medical staff (no data entry from interviewees/patients) [56-73]
- Study duration less than one month or one-time data entry [47,74-100]
- No possibilities, problems, or perspectives mentioned in the study [101]
- Reviews dealing with smartphone apps, mobile phone surveys, and new data-collection methods in general [102-105]

Thus, studies included in the review met the following criteria:

- Data collection with a smartphone app, including apps used for disease screening
- Apps either prescribed by a physician or selected by the user
- More than self-management functions
- App available on smartphone
- Data entry completed by the interviewees/patients (or close relatives)
- Repeated data entry on a longitudinal basis (more than one month)
- Possibilities, problems, and perspectives of data-collection apps described
- Feasibility studies with attitudes toward sharing health information (with researchers or physicians) within smartphone apps or smartphone devices

The final summary consisted of 17 studies, which were summarized using content analysis (Figure 1). There were 3 guiding questions for the content analysis: (1) in which settings have apps for data collection been used so far, (2) what challenges and requirements exist regarding the implementation of apps, and (3) what potential does data collection with smartphone devices have? These aspects are synthesized in the Results based on those aspects described in the primary articles.

Figure 1. Flowchart of the scoping review.



Results

Overview: Studies Collecting Patient Health Data Over Time

A total of 17 studies were included in the synthesis of the scoping review (Table 1). These studies were conducted for the following reasons:

- To track real-time changes in symptom(s) (severity) or other disease-related patterns (for creating modified/redefined treatment plans) [106-115]
- To determine users’ compliance and engagement with technology adoption, patient data collection, and provider communication with smartphone app [107,111,112,114,115]
- To test an app for aftercare assessment [116-118]
- To screen for certain (disease or patient-related) outcomes [119]
- To develop an app-based participatory surveillance system for collecting syndromic data [120]
- To provide a practical guide to developing and implementing a longitudinal study with an app [121]

Of these 17 studies, 7 were feasibility or pilot studies. The country with the largest number of studies was the US (7/17); 2 studies based in Sweden and 2 more were based in the Netherlands. All (feasibility/pilot and full-scale) studies on data collection over a longer period of time using smartphone devices used a medical or health-related indication, with a wide range of indication areas. The two main indication areas were psychosomatic disorders (eg, depression, mood, post-traumatic stress disorder) and neurodegenerative diseases (eg, Parkinson Disease). The other studies focused on indication areas such as respiratory diseases (eg, asthma), cardiovascular diseases (eg, acute stroke, atrial fibrillation), malignant neoplasms (eg, prostate cancer), sleeping disorders (eg, insomnia, sleep apnea), infectious diseases (eg, acute febrile illness and gastroenteritis), weight loss in obese people, gestational diabetes mellitus, abdominal surgery, chronic pain, and drinking behavior in adolescents. Target groups were related to pregnancy (breastfeeding [108], depression [109], and gestational diabetes mellitus [122]) or focused on children [120] and young people [119].

Table 1. Overview of studies included in the scoping review.

Reference	Feasibility or pilot study	Country	Indication area	Study design	Duration of data collection	Participants at baseline, n	Specific target group
Bot et al (2016) [106]	✓	US	Parkinson disease	Observational	6 months	8320	N/A[a]
Burke et al (2017) [121]	✓	US	Obesity	Observational	12 months	151	N/A
Chan et al (2017) [107]	X[b]	US	Asthma	Observational	6 months	7593	N/A
Cooray et al (2015) [116]	X	Sweden	Stroke	Cohort	2 months	48	N/A
Demirci and Bogen (2017) [108]	X	US	Breastfeeding	Observational	8 weeks	61	Pregnancy
Faherty et al (2017) [109]	X	US	Depression	Cohort	8 weeks	36	Pregnancy
Horsch et al (2017) [110]	X	Netherlands	Insomnia	RCT[c]	6-7 weeks (with 3-month follow-up)	151	N/A
Isetta et al (2017) [111]	✓	Spain	Obstructive sleep apnea	Observational	6 weeks	60	N/A
Jamison et al (2017) [112]	✓	US	Chronic pain	Observational	3 months (with option for 6 months)	105	N/A
Labhart et al (2017) [119]	X	Switzerland	Alcohol consumption	Observational	7 weeks	176	Young people
Noe et al (2017) [113]	X	Wales	Mood tracking	Observational	8 weeks	76	N/A
Olson et al (2017) [120]	X	Guatemala	Acute febrile illness and acute gastroenteritis	Cohort	9 months	469	Children
Pavliscsak et al (2016) [117]	✓	England	PTSD and/or TBI in rehabilitation	RCT	36 weeks	95	N/A
Peleg et al (2017) [122]	✓	Italy and Spain	Atrial fibrillation and gestational diabetes mellitus	Cohort	9 months	29	Pregnancy
Silva de Lima et al (2017) [114]	X	Netherlands and North America	Parkinson disease	Observational	6-13 weeks	953	N/A
Sundberg et al (2017) [115]	X	Sweden	Prostate cancer	Intervention	11 weeks	130	N/A
Symer et al (2017) [118]	✓	US	Abdominal surgery	Observational	4 weeks	31	N/A

[a] N/A: not applicable.

[b] Studies marked with an “X” are not feasibility or pilot studies.

[c] RCT: randomized controlled trial.

The sample sizes differed greatly, ranging from 29 to 8320 study participants at baseline. The participants’ mean ages ranged from 7 years [120] (although children did not enter data unsupervised, and so parents were necessarily involved) to 69 years [115], although not every study provided information on the (mean) age of study participants. The majority of studies (10/17) had an observational design, with one using an ecological momentary assessment [121]. Of the observational studies, 4 were cohort studies, 2 were randomized controlled trials, and 1 was a nonrandomized controlled intervention trial. The duration of data collection varied between the studies from 4 weeks to 12 months. Further details are provided in Multimedia Appendix 1.

Potential of Using Smartphone Devices for Data Collection Over Time

Within the reviewed articles, several potential benefits of using smartphone devices for data collection are described. Firstly, automatic push notifications offer the potential to provide daily, weekly, or monthly reminders or any kind of information [117]. This can be linked to multiple further functions, showing advantages for researchers as well as for users. For that reason, an app can be used as a data-collection tool or as a reminder for the elderly, has potential for self-interventions (eg, when users can track their symptoms), and can be used for real-time information transfer. Patients tracking their symptoms, for example, can use apps to help themselves monitor their diseases, as well as help physicians to create tailor-made treatment plans [117]. Some other advantages can be derived from sharing real-time data between researchers via (secure) platforms as well as storing and replacing data from one software function to another without data loss or mistakes in data entry [106,116]. Rapid data synchronization between data entry (from the user) and demand-oriented provision of the data (for the researcher or the physician) is also possible [116]. Data entry occurs almost simultaneously with its storage and access, which saves time and increases effectivity. With the support of smartphone devices, data collection is possible on a daily basis over long periods of time, and questionnaire responses can be tracked longitudinally [117]. Smartphone-based interventions, or aftercare assessment via the smartphone (in lieu of going to a hospital or seeing a physician), remove barriers and save time; they also remove administrative barriers and travel difficulties (eg, for aftercare assessment). Furthermore, data loss due to reminder bias can be prevented [116].

With sensitive (health) data, there is always a need for data protection. When using secure data storage and unique random codes for user identification, anonymity and data security can be ensured by using apps [106,117].

Challenges in Using Smartphone Devices for Data Collection Over Time

Although data collection with smartphone apps has great potential, there are also challenges and disadvantages. As is the case for traditional methods of data collection, follow-up with smartphones can be nonuniform [106]. For that reason, users' (or patients') engagement and compliance over longer periods of time is a critical issue. Optimally, for conducting studies on smartphone devices, people need to own smartphones with an available mobile internet connection. Otherwise, it is necessary to provide these devices to the participants. This might be feasible for a cross-sectional study of short duration, but for large sample sizes and over longer periods of time this is an obstacle [110]. Furthermore, participant-related preferences are to carry no more than one device [109,121]. Overall, a risk of loss to follow-up exists [108], as in other longitudinal data collections, which might be due to either lack of motivation or technical issues, such as poor data signal or participant discomfort using mobile phones [120]. Therefore, interactive and feedback elements and rewards may lead to greater adherence [112,113].

Discussion

Overview

Although the studies identified in this scoping review claimed significant potential for health-related data collection using smartphone devices, it is obvious that very few approaches have integrated these devices into clinical or epidemiological research activities so far. This is particularly true for the data collection in longitudinal study designs. Although apps can be simply tailored to the needs of target groups [12], only a few studies [108,109,120,122] focused explicitly on specified groups. Most of the articles identified in this review dealt with a particular medical or health-related issue. However, apart from pregnancy and childhood, the study participants consisted of nonspecific target groups. For example, we were not able to identify any study particularly focusing on migrant populations. Therefore, the main result of this overview about mobile apps for data collection in epidemiological research is that further studies and, depending on the results of such studies, adequate solutions for implementing modern technologies for data collection in a priori defined target groups is urgently needed. This is also a requirement for achieving high usability. Although it is possible to design an app using free or low-cost development tools, this can be difficult when app designers are unfamiliar with these tools. In such cases, a pilot test with a subpopulation is highly recommended [11]. This is particularly important because a previous study [10] has indicated greater commitment and engagement in using app-based data collection tools, especially among young and middle-aged populations as compared to older populations.

General population surveys frequently do not adequately represent population subgroups. This challenge is reinforced by the fact that specific groups, such as migrants or refugees, are difficult-to-sample populations. These challenges apply to the fact that such difficult-to-sample populations can be rare, difficult to locate, difficult to enumerate, and/or difficult to interview [123]. Therefore, the implementation of mobile apps for data collection might be particularly suitable for vulnerable but also highly mobile populations, such as migrants, in general, or refugees in particular. Accurate data about health status, health determinants, use of health-care services, and risk behaviors among migrants are needed in order to monitor health and improve health services among this population subgroup [124]. As of yet, health-related data on migrants, and refugees in particular, are scarce, and data-collection approaches face several methodological limitations.

As already highlighted in a narrative review of data-collection practices for monitoring migrant health in Europe [124], there are several legal obstacles related to data protection which need to be taken into account. Due to the recently enforced European General Data Protection Regulation [125], aspects of informed consent, privacy and confidentiality, and the right to withdrawal or omit items will receive further attention in the future. These new data protection regulations need to be kept in mind when developing and implementing data-collection tools. These tools should provide additional information to already established instruments such as the census, death registers, and disease surveillance systems. These established instruments only partially include information on migration status, and if they do, the information is only available at a highly aggregated level [126]. Further information on the health of migrants is available based on epidemiological studies of population samples using medical diagnoses [127]. However, these analyses face the challenge that migrants may encounter barriers in accessing health services. This may lead to an underestimation of the disease burden [127].

To overcome these challenges, which lead to an underrepresentation of specific subgroups and, therefore, insufficient data for action [128], several measures have already been taken. These measures include, but are not limited to, disproportionately allocated sampling, multiplicity sampling, and the use of multiple frames. However, it has to be noted that oversampling of people with a migration background is not sufficient to avoid systematic bias in the sample due to nonparticipation. Further measures, such as personal contacting, multilingual instruments and interviewers, and extensive public relations, have to be taken into account. All of this can be done within smartphone-based apps. For the target groups of migrants in epidemiological studies, no singe simple solution exists [123,129,130]. However, studies indicate that constant communication and follow-up of study participants enhances the response rates in longitudinal studies and improves the quality of data [131]. For that reason, mobile apps may provide solutions for overcoming common challenges in epidemiological data collection.

Limitations

Overall, the results of this scoping review need to be interpreted with caution. We were only able to provide a broad overview of challenges and potential benefits linked with data collection using mobile apps described in the literature. The scoping review was not focused on a specific topic in epidemiological research, in order to capture all the ideas and experiences previously described in the literature. Due to the comparatively broad (but not concrete) search algorithm that included terms related to mobile apps in combination with study designs, we might have missed some studies which have applied mobile technologies in epidemiological data collection. For example, mobile phone surveys have been excluded. However, although this scoping review was conducted in only one literature database (PubMed), we anticipate that it will provide a comprehensive overview of the potential benefits and challenges that researchers may face when using mobile apps for data collection in epidemiological studies. We were not able to identify studies specifically dealing with the target group of migrants or refugees.

Conclusions

It appears that the challenges connected to data collection among migrant subgroups can be overcome by using the innovative technologies of mobile apps. These approaches allow the recruitment of study participants from diverse ethnic backgrounds when designed in a migration-sensitive, cross-language, and cross-cultural fashion. Continuing contact with study participants can be ensured during longitudinal studies. This is of particular relevance for mobile populations, such as refugees, who might not have reached their place of destination at the time of recruitment. Furthermore, due to the cost-effectiveness of app-based data collection, the challenges of small sample sizes and low response rates can be at least partially overcome. However, feasibility studies need to be conducted in the near future to test the applicability and acceptance of mobile apps for epidemiological research among various subpopulations.

Acknowledgments

We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité – Universitätsmedizin Berlin.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary appendix.
[PDF File (Adobe PDF File), 96 KB - jmir_v23i1e17691_app1.pdf]

References

1. Statista. Anzahl der Smartphone-Nutzer weltweit von 2016 bis 2019 und Prognose bis 2023. 2021. URL: https://de.statista.com/statistik/daten/studie/309656/umfrage/prognose-zur-anzahl-der-smartphone-nutzer-weltweit/ [accessed 2021-01-14]
2. Research 2 Guidance. mHealth App Developer Economics 2015: The current status and trends of the mHealth app market. 2017. URL: http://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2015.pdf [accessed 2021-01-14]
3. Kim JH, Kwon S, Shim SR, Sun HY, Ko YM, Chun D, et al. Validation and reliability of a smartphone application for the International Prostate Symptom Score questionnaire: a randomized repeated measures crossover study. J Med Internet Res 2014 Feb 10;16(2):e38 [FREE Full text] [doi: 10.2196/jmir.3042] [Medline: 24513507]
4. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]

5. Behar J, Roebuck A, Domingos JS, Gederi E, Clifford GD. A review of current sleep screening applications for smartphones. Physiol Meas 2013 Jun 17;34(7):R29-R46. [doi: 10.1088/0967-3334/34/7/r29]
6. Pfaeffli L, Maddison R, Jiang Y, Dalleck L, Löf M. Measuring Physical Activity in a Cardiac Rehabilitation Population Using a Smartphone-Based Questionnaire. J Med Internet Res 2013 Mar 22;15(3):e61. [doi: 10.2196/jmir.2419]
7. Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. BMC Psychiatry 2012 Oct 17;12:172 [FREE Full text] [doi: 10.1186/1471-244X-12-172]
8. Dufau S, Duñabeitia JA, Moret-Tatay C, McGonigal A, Peeters D, Alario F, et al. Smart phone, smart science: how the use of smartphones can revolutionize research in cognitive science. PLoS One 2011 Sep;6(9):e24974 [FREE Full text] [doi: 10.1371/journal.pone.0024974] [Medline: 21980370]
9. Abelson JS, Symer M, Peters A, Charlson M, Yeo H. Mobile health apps and recovery after surgery: What are patients willing to do? The American Journal of Surgery 2017 Oct;214(4):616-622. [doi: 10.1016/j.amjsurg.2017.06.009]
10. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2014 Oct 17;22(1):29-34. [doi: 10.1136/amiajnl-2014-002840]
11. Burnay E, Cruz-Correia R, Jacinto T, Sousa AS, Fonseca J. Challenges of a mobile application for asthma and allergic rhinitis patient enablement-interface and synchronization. Telemed J E Health 2013 Jan;19(1):13-18. [doi: 10.1089/tmj.2012.0020] [Medline: 23215639]
12. Hartin P, Nugent C, McClean S. A smartphone application to evaluate technology adoption and usage in persons with dementia. 2014 Presented at: Conf Proc IEEE Eng Med Biol Soc; 2014; Chicago p. 5389-5392. [doi: 10.1109/embc.2014.6944844]
13. Min YH, Lee JW, Shin Y, Jo M, Sohn G, Lee J, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014 May;16(5):e135 [FREE Full text] [doi: 10.2196/jmir.3421] [Medline: 24860070]
14. Mulhern B, O’Gorman H, Rotherham N, Brazier J. Comparing the measurement equivalence of EQ-5D-5L across different modes of administration. Health Qual Life Outcomes 2015 Nov 26;13:191. [doi: 10.1186/s12955-015-0382-6]
15. Serrano KJ, Yu M, Riley WT, Patel V, Hughes P, Marchesini K, et al. Willingness to Exchange Health Information via Mobile Devices: Findings From a Population-Based Survey. Ann Fam Med 2016 Jan;14(1):34-40 [FREE Full text] [doi: 10.1370/afm.1888] [Medline: 26755781]
16. Wang J, Yao NA, Liu Y, Geng Z, Wang Y, Shen N, et al. Development of a Smartphone Application to Monitor Pediatric Patient-Reported Outcomes. Stud Health Technol Inform 2017;245:253-257. [Medline: 29295093]
17. Marcano Belisario JS, Doherty K, O'Donoghue J, Ramchandani P, Majeed A, Doherty G, et al. A bespoke mobile application for the longitudinal assessment of depression and mood during pregnancy: protocol of a feasibility study. BMJ Open 2017 May 29;7(5):e014469. [doi: 10.1136/bmjopen-2016-014469]
18. Abu-Ghanem S, Handzel O, Ness L, Ben-Artzi-Blima M, Fait-Ghelbendorf K, Himmelfarb M. Smartphone-based audiometric test for screening hearing loss in the elderly. Eur Arch Otorhinolaryngol 2016;273(2):333-339. [doi: 10.1007/s00405-015-3533-9] [Medline: 25655259]
19. Brooks GC, Vittinghoff E, Iyer S, Tandon D, Kuhar P, Madsen KA, et al. Accuracy and Usability of a Self-Administered 6-Minute Walk Test Smartphone Application. Circ Heart Fail 2015 Sep;8(5):905-913. [doi: 10.1161/CIRCHEARTFAILURE.115.002062]
20. Dempster NJ, Risk R, Clark R, Meddings RN. Urologists' usage and perceptions of urological apps. J Telemed Telecare 2014 Dec;20(8):450-453. [doi: 10.1177/1357633X14555622] [Medline: 25316043]
21. Sheehan B, Lee Y, Rodriguez M, Tiase V, Schnall R. A comparison of usability factors of four mobile devices for accessing healthcare information by adolescents. Appl Clin Inform 2012 Dec;3(4):356-366 [FREE Full text] [doi: 10.4338/ACI-2012-06-RA-0021] [Medline: 23227134]
22. Snipes SA, Smyth JM, Murphy D, Miranda PY, Ishino FAM. Provision Increases Reported PPE Use for Mexican Immigrant Farmworkers. Journal of Occupational and Environmental Medicine 2015;57(12):1343-1346. [doi: 10.1097/jom.0000000000000563]
23. Tawara S, Yonemochi Y, Kosaka T, Kouzaki Y, Takita T, Tsuruta T. Use of Patients' Mobile Phones to Store and Share Personal Health Information: Results of a Questionnaire Survey. Intern Med 2013;52(7):751-756. [doi: 10.2169/internalmedicine.52.9030]
24. Zhang MW, Ho CS, Fang P, Lu Y, Ho RC. Methodology of developing a smartphone application for crisis research and its clinical application. Technol Health Care 2014 Aug 01;22(4):547-559. [doi: 10.3233/thc-140819]
25. Bajaj JS, Thacker LR, Heuman DM, Fuchs M, Sterling RK, Sanyal AJ, et al. The Stroop smartphone application is a short and valid method to screen for minimal hepatic encephalopathy. Hepatology 2013 Sep;58(3):1122-1132 [FREE Full text] [doi: 10.1002/hep.26309] [Medline: 23389962]
26. Mehta S, Barker K, Bowman B, Galloway H, Oliashirazi N, Oliashirazi A. Reliability, Concurrent Validity, and Minimal Detectable Change for iPhone Goniometer App in Assessing Knee Range of Motion. J Knee Surg 2016 Nov 28;30(6):577-584. [doi: 10.1055/s-0036-1593877]

27. Arsand E, Tatara N, ostengen G, Hartvigsen G. Mobile Phone-Based Self-Management Tools for Type 2 Diabetes: The Few Touch Application. Journal of Diabetes Science and Technology 2010 Mar 01;4(2):328-336. [doi: 10.1177/193229681000400213]
28. Azevedo R, Bernardes M, Fonseca J, Lima A. Smartphone application for rheumatoid arthritis self-management: cross-sectional study revealed the usefulness, willingness to use and patients' needs. Rheumatol Int 2015;35(10):1675-1685. [doi: 10.1007/s00296-015-3270-9] [Medline: 25903352]
29. Becker S, Brandl C, Meister S, Nagel E, Miron-Shatz T, Mitchell A, et al. Demographic and health related data of users of a mobile application to support drug adherence is associated with usage duration and intensity. PLoS One 2015 Jan;10(1):e0116980 [FREE Full text] [doi: 10.1371/journal.pone.0116980] [Medline: 25629939]
30. Becker S, Kribben A, Meister S, Diamantidis CJ, Unger N, Mitchell A. User Profiles of a Smartphone Application to Support Drug Adherence — Experiences from the iNephro Project. PLoS ONE 2013 Oct 23;8(10):e78547. [doi: 10.1371/journal.pone.0078547]
31. Liu X, Wang R, Zhou D, Hong Z. Feasibility and acceptability of smartphone applications for seizure self-management in China: Questionnaire study among people with epilepsy. Epilepsy Behav 2016 Feb;55:57-61. [doi: 10.1016/j.yebeh.2015.11.024] [Medline: 26745631]
32. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909 [FREE Full text] [doi: 10.1111/pme.12327] [Medline: 24422990]
33. Shah N, Jonassaint J, De Castro L. Patients Welcome the Sickle Cell Disease Mobile Application to Record Symptoms Technology (SMART). Hemoglobin 2014 Feb 10;38(2):99-103. [doi: 10.3109/03630269.2014.880716]
34. Tregarthen JP, Lock J, Darcy AM. Development of a smartphone application for eating disorder self-monitoring. Int J Eat Disord 2015 Jul 27;48(7):972-982. [doi: 10.1002/eat.22386]
35. Zhang MW, Ho RC, Hawa R, Sockalingam S. Pilot implementation and user preferences of a Bariatric After-care application. Technol Health Care 2015 Oct 27;23(6):729-736. [doi: 10.3233/thc-151025]
36. Atkinson KM, Ducharme R, Westeinde J, Wilson SE, Deeks SL, Pascali D, et al. Vaccination attitudes and mobile readiness: A survey of expectant and new mothers. Human Vaccines & Immunotherapeutics 2015 Feb 25;11(4):1039-1045. [doi: 10.1080/21645515.2015.1009807]
37. Conway N, Campbell I, Forbes P, Cunningham S, Wake D. mHealth applications for diabetes: User preference and implications for app development. Health Informatics J 2016;22(4):1111-1120. [doi: 10.1177/1460458215616265] [Medline: 26635324]
38. Alenazi H, Alghamdi M, Alradhi S, Househ M, Zakaria N. A Study on Saudi Diabetic Patients' Readiness to Use Mobile Health. Stud Health Technol Inform 2017;245:1210. [Medline: 29295297]
39. Swendeman D, Ramanathan N, Baetscher L, Medich M, Scheffler A, Comulada WS, et al. Smartphone Self-Monitoring to Support Self-Management Among People Living With HIV. J Acquir Immune Defic Syndr 2015 May 01;69(Suppl. 1):S80-S91. [doi: 10.1097/qai.0000000000000570]
40. Barrio P, Ortega L, López H, Gual A. Self-management and Shared Decision-Making in Alcohol Dependence via a Mobile App: a Pilot Study. Int J Behav Med 2017 Feb 24;24(5):722-727. [doi: 10.1007/s12529-017-9643-6]
41. Con D, Jackson B, Gray K, De Cruz P. eHealth for inflammatory bowel disease self-management – the patient perspective. Scand J Gastroenterol 2017 Jun 9;52(9):973-980. [doi: 10.1080/00365521.2017.1333625]
42. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: A self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan;64(1):139-145. [doi: 10.1002/pbc.26177] [Medline: 27574031]
43. Eisenhauer CM, Hageman PA, Rowland S, Becker BJ, Barnason SA, Pullen CH. Acceptability of mHealth Technology for Self-Monitoring Eating and Activity among Rural Men. Public Health Nurs 2017 Mar;34(2):138-146. [doi: 10.1111/phn.12297] [Medline: 27757986]
44. Lee J, Song S, Ahn J, Kim Y, Lee J. Use of a Mobile Application for Self-Monitoring Dietary Intake: Feasibility Test and an Intervention Study. Nutrients 2017 Jul 13;9(7):748. [doi: 10.3390/nu9070748]
45. Simpson AJ, Honkoop PJ, Kennington E, Snoeck-Stroband J, Smith I, East J, et al. Perspectives of patients and healthcare professionals on mHealth for asthma self-management. Eur Respir J 2017 May 01;49(5):1601966. [doi: 10.1183/13993003.01966-2016]
46. Sommer J, Daus M, Smith M, Luna D. Mobile Application for Pregnant Women: What Do Mothers Say? Stud Health Technol Inform 2017;245:221-224. [Medline: 29295086]
47. Yu DX, Parmanto B, Dicianno BE, Watzlaf VJ, Seelman KD. Accessibility needs and challenges of a mHealth system for patients with dexterity impairments. Disabil Rehabi Assist Technol 2017 Jan;12(1):56-64. [doi: 10.3109/17483107.2015.1063171]
48. Jiam NT, Hoon AH, Hostetter CF, Khare MM. IIAM (important information about me): a patient portability profile app for adults, children and families with neurodevelopmental disabilities. Disabil Rehabil Assist Technol 2017 Dec;12(6):599-604. [doi: 10.1080/17483107.2016.1198435] [Medline: 27377240]

49. Bielli E, Carminati F, La CS, Lina M, Brunelli C, Tamburini M. A Wireless Health Outcomes Monitoring System (WHOMS): development and field testing with cancer patients using mobile phones. BMC Med Inform Decis Mak 2004 Jun 15;4:7 [FREE Full text] [doi: 10.1186/1472-6947-4-7] [Medline: 15196308]
50. Dunton GF, Liao Y, Intille SS, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011 Jun;19(6):1205-1212 [FREE Full text] [doi: 10.1038/oby.2010.302] [Medline: 21164502]
51. Eagle N, Pentland A, Lazer D. Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences 2009 Aug 17;106(36):15274-15278. [doi: 10.1073/pnas.0900282106]
52. Jovanov E, Frith K, Anderson F, Milosevic M, Shrove M. Real-time monitoring of occupational stress of nurses. 2011 Presented at: Conf Proc IEEE Eng Med Biol Soc; 2011; Boston p. 3640-3643. [doi: 10.1109/iembs.2011.6090612]
53. Dewulf B, Neutens T, Lefebvre W, Seynaeve G, Vanpoucke C, Beckx C, et al. Dynamic assessment of exposure to air pollution using mobile phone data. Int J Health Geogr 2016 Apr 21;15(1):14. [doi: 10.1186/s12942-016-0042-z]
54. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
55. McLean A, Osgood N, Newstead-Angel J, Stanley K, Knowles D, van der Kamp W, et al. Building Research Capacity: Results of a Feasibility Study Using a Novel mHealth Epidemiological Data Collection System Within a Gestational Diabetes Population. Stud Health Technol Inform 2017;234:228-232. [Medline: 28186046]
56. Katib A, Rao D, Rao P, Williams K, Grant J. A prototype of a novel cell phone application for tracking the vaccination coverage of children in rural communities. Comput Methods Programs Biomed 2015 Nov;122(2):215-228. [doi: 10.1016/j.cmpb.2015.08.008]
57. Haskew J, Kenyi V, William J, Alum R, Puri A, Mostafa Y, et al. Use of Mobile Information Technology during Planning, Implementation and Evaluation of a Polio Campaign in South Sudan. PLoS ONE 2015 Aug 7;10(8):e0135362. [doi: 10.1371/journal.pone.0135362]
58. Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an Android-based mHealth system for population surveillance in developing countries. J Am Med Inform Assoc 2012 Jul 01;19(4):655-659. [doi: 10.1136/amiajnl-2011-000476]
59. Rotheram-Borus M, Richter L, Van RH, van HA, Tomlinson M, Stein A, et al. Project Masihambisane: a cluster randomised controlled trial with peer mentors to improve outcomes for pregnant mothers living with HIV. Trials 2011 Jan 04;12:2 [FREE Full text] [doi: 10.1186/1745-6215-12-2] [Medline: 21205302]
60. Setswe G, Muyanga S, Witthun J, Nyasulu P. Public awareness and knowledge of the National Health Insurance in South Africa. Pan Afr Med J 2015;22(19):1-10. [doi: 10.11604/pamj.2015.22.19.6131]
61. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. BMC Med Inform Decis Mak 2009 Dec 23;9:51 [FREE Full text] [doi: 10.1186/1472-6947-9-51] [Medline: 20030813]
62. Vélez O, Okyere PB, Kanter AS, Bakken S. A usability study of a mobile health application for rural Ghanaian midwives. J Midwifery Womens Health 2014 Jan;59(2):184-191 [FREE Full text] [doi: 10.1111/jmwh.12071] [Medline: 24400748]
63. Zhang S, Wu Q, van Velthoven MH, Chen L, Car J, Rudan I, et al. Smartphone Versus Pen-and-Paper Data Collection of Infant Feeding Practices in Rural China. J Med Internet Res 2012 Sep 18;14(5):e119. [doi: 10.2196/jmir.2183]
64. Knoble SJ, Bhusal MR. Electronic diagnostic algorithms to assist mid-level health care workers in Nepal: A mixed-method exploratory study. Int J Med Inform 2015 May;84(5):334-340. [doi: 10.1016/j.ijmedinf.2015.01.011]
65. Andrew BY, Stack CM, Yang JP, Dodds JA. mStroke: "Mobile Stroke"—Improving Acute Stroke Care with Smartphone Technology. J Stroke Cerebrovasc Dis 2017 Jul;26(7):1449-1456. [doi: 10.1016/j.jstrokecerebrovasdis.2017.03.016]
66. Bakibinga P, Kamande E, Omuya M, Ziraba AK, Kyobutungi C. The role of a decision-support smartphone application in enhancing community health volunteers' effectiveness to improve maternal and newborn outcomes in Nairobi, Kenya: quasi-experimental research protocol. BMJ Open 2017 Jul 20;7(7):e014896 [FREE Full text] [doi: 10.1136/bmjopen-2016-014896] [Medline: 28729309]
67. Coppock D, Zambo D, Moyo D, Tanthuma G, Chapman J, Re VL, et al. Development and Usability of a Smartphone Application for Tracking Antiretroviral Medication Refill Data for Human Immunodeficiency Virus. Methods Inf Med 2018 Jan 24;56(5):351-359. [doi: 10.3414/me17-01-0045]
68. Kenny A, Gordon N, Griffiths T, Kraemer JD, Siedner MJ. Validation Relaxation: A Quality Assurance Strategy for Electronic Data Collection. J Med Internet Res 2017 Aug 18;19(8):e297. [doi: 10.2196/jmir.7813]
69. Mishori R, Anastario M, Naimer K, Varanasi S, Ferdowsian H, Abel D, et al. mJustice: Preliminary Development of a Mobile App for Medical-Forensic Documentation of Sexual Violence in Low-Resource Environments and Conflict Zones. Glob Health Sci Pract 2017 Mar 28;5(1):138-151. [doi: 10.9745/ghsp-d-16-00233]
70. Modi D, Desai S, Dave K, Shah S, Desai G, Dholakia N, et al. Cluster randomized trial of a mHealth intervention "ImTeCHO" to improve delivery of proven maternal, neonatal, and child care interventions through community-based Accredited Social Health Activists (ASHAs) by enhancing their motivation and strengthening supervision in tribal areas of Gujarat, India: study protocol for a randomized controlled trial. Trials 2017 Jun 9;18(1):270. [doi: 10.1186/s13063-017-1998-0]

71. Motulsky A, Wong J, Cordeau J, Pomalaza J, Barkun J, Tamblyn R. Using mobile devices for inpatient rounding and handoffs: an innovative application developed and rapidly adopted by clinicians in a pediatric hospital. J Am Med Inform Assoc 2016 Aug 22;24(e1):e69-e78. [doi: 10.1093/jamia/ocw107]
72. Vedachalam S, MacDonald LH, Shiferaw S, Seme A, Schwab KJ. Underreporting of high-risk water and sanitation practices undermines progress on global targets. PLoS ONE 2017 May 10;12(5):e0176272. [doi: 10.1371/journal.pone.0176272]
73. Zhou L, Watzlaf V, Abernathy P, Abdelhak M. A Health Information System for Scalable and Comprehensive Assessment of Well-Being: A Multidisciplinary Team Solution. Perspect Health Inf Manag 2017 Jul 01;14(Summer):1d [FREE Full text] [Medline: 28855857]
74. Runyan JD, Steenbergh TA, Bainbridge C, Daugherty DA, Oke L, Fry BN. A smartphone ecological momentary assessment/intervention. PLoS One 2013 Aug;8(8):e71325 [FREE Full text] [doi: 10.1371/journal.pone.0071325] [Medline: 23977016]
75. Spook JE, Paulussen T, Kok G, Van EP. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
76. Webb JR, Webb BF, Schroeder MC, North CS. Association of aphthous ulcers with self-reported symptoms of depression in a sample of smartphone users. Ann Clin Psychiatry 2013 Nov;25(4):266-270. [Medline: 24199216]
77. Seto E, Hua J, Wu L, Shia V, Eom S, Wang M, et al. Models of Individual Dietary Behavior Based on Smartphone Data: The Influence of Routine, Physical Activity, Emotion, and Food Environment. PLoS One 2016;11(4):e0153085 [FREE Full text] [doi: 10.1371/journal.pone.0153085] [Medline: 27049852]
78. Reid SC, Kauer SD, Dudgeon P, Sanci LA, Shrier LA, Patton GC. A mobile phone program to track young people's experiences of mood, stress and coping. Development and testing of the mobiletype program. Soc Psychiatry Psychiatr Epidemiol 2009 Jun;44(6):501-507. [doi: 10.1007/s00127-008-0455-5] [Medline: 19011723]
79. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Construct validity and reliability of a real-time multidimensional smartphone app to assess pain in children and adolescents with cancer. Pain 2015 Dec;156(12):2607-2615. [doi: 10.1097/j.pain.0000000000000385] [Medline: 26580680]
80. Brannon EE, Cushing CC, Crick CJ, Mitchell TB. The promise of wearable sensors and ecological momentary assessment measures for dynamical systems modeling in adolescents: a feasibility and acceptability study. Transl Behav Med 2016 Dec;6(4):558-565 [FREE Full text] [doi: 10.1007/s13142-016-0442-4] [Medline: 27678501]
81. Warren-Stomberg M, Jacobsson J, Brattwall M, Jildenstål P. At-home monitoring after surgery/anaesthesia - a challenge. J Eval Clin Pract 2016 May 01;22(6):886-890. [doi: 10.1111/jep.12551]
82. Alqahtani AS, Rashid H, Basyouni MH, Alhawassi TM, BinDhim NF. Public response to MERS-CoV in the Middle East: iPhone survey in six countries. J Infect Public Health 2017 Sep;10(5):534-540. [doi: 10.1016/j.jiph.2016.11.015]
83. Chen Y, Wong J, Ayob A, Othman N, Poh B. Can Malaysian Young Adults Report Dietary Intake Using a Food Diary Mobile Application? A Pilot Study on Acceptability and Compliance. Nutrients 2017 Jan 13;9(1):62. [doi: 10.3390/nu9010062]
84. Farnham A, Furrer R, Blanke U, Stone E, Hatz C, Puhan M. The quantified self during travel: mapping health in a prospective cohort of travellers. J Travel Med 2017 Sep 01;24(5):tax050. [doi: 10.1093/jtm/tax050] [Medline: 28931147]
85. Fernández-Castro J, Martínez-Zaragoza F, Rovira T, Edo S, Solanes-Puchol ?, Martín-del-Río B, et al. How does emotional exhaustion influence work stress? Relationships between stressor appraisals, hedonic tone, and fatigue in nurses’ daily tasks: A longitudinal cohort study. Int J Nurs Stud 2017 Oct;75:43-50. [doi: 10.1016/j.ijnurstu.2017.07.002]
86. Gomes MS, Bonan PRF, Ferreira VYN, de Lucena Pereira L, Correia RJC, da Silva Teixeira HB, et al. Development of a mobile application for oral cancer screening. Technol Health Care 2017 Mar 22;25(2):187-195. [doi: 10.3233/thc-161259]
87. Hennig T, Krkovic K, Lincoln TM. What predicts inattention in adolescents? An experience-sampling study comparing chronotype, subjective, and objective sleep parameters. Sleep Med 2017 Oct;38:58-63. [doi: 10.1016/j.sleep.2017.07.009]
88. Kaplan-Neeman R, Muchnik C, Amir N. Listening to music with personal listening devices: monitoring the noise dose using a smartphone application. Int J Audiol 2017 Mar 10;56(6):400-407. [doi: 10.1080/14992027.2017.1297541]
89. Liddle J, Wishink A, Springfield L, Gustafsson L, Ireland D, Silburn P. Can smartphones measure momentary quality of life and participation? A proof of concept using experience sampling surveys with university students. Aust Occup Ther J 2017 Apr 06;64(4):294-304. [doi: 10.1111/1440-1630.12360]
90. Mindell JA, Leichman ES, Walters RM. Sleep location and parent-perceived sleep outcomes in older infants. Sleep Med 2017 Nov;39:1-7. [doi: 10.1016/j.sleep.2017.08.003]
91. Moore RC, Kaufmann CN, Rooney AS, Moore DJ, Eyler LT, Granholm E, et al. Feasibility and Acceptability of Ecological Momentary Assessment of Daily Functioning Among Older Adults with HIV. Am J Geriatr Psychiatry 2017 Aug;25(8):829-840. [doi: 10.1016/j.jagp.2016.11.019]
92. Moran EK, Culbreth AJ, Barch DM. Ecological momentary assessment of negative symptoms in schizophrenia: Relationships to effort-based decision making and reinforcement learning. J Abnorm Psychol 2017 Jan;126(1):96-105. [doi: 10.1037/abn0000240]

93. Roberts ME, Lu B, Browning CR, Ferketich AK. Tracking Young Adults' Attitudes Toward Tobacco Marketing Using Ecological Momentary Assessment (EMA). Subst Use Misuse 2017 Jun 12;52(9):1208-1213. [doi: 10.1080/10826084.2017.1302958]
94. Roy R, Rangan A, Hebden L, Yu Louie JC, Tang LM, Kay J, et al. Dietary contribution of foods and beverages sold within a university campus and its effect on diet quality of young adults. Nutrition 2017 Feb;34:118-123. [doi: 10.1016/j.nut.2016.09.013]
95. Sperry SH, Kwapil TR. What can daily life assessment tell us about the bipolar spectrum? Psychiatry Res 2017 Jun;252:51-56. [doi: 10.1016/j.psychres.2017.02.045]
96. Timmer BHB, Hickson L, Launer S. Ecological Momentary Assessment: Feasibility, Construct Validity, and Future Applications. Am J Audiol 2017 Oct 12;26(3S):436-442. [doi: 10.1044/2017_aja-16-0126]
97. van Wel L, Huss A, Bachmann P, Zahner M, Kromhout H, Fröhlich J, et al. Context-sensitive ecological momentary assessments; integrating real-time exposure measurements, data-analytics and health assessment using a smartphone application. Environ Int 2017 Jun;103:8-12. [doi: 10.1016/j.envint.2017.03.016]
98. Wang J, Wang Q, Wimalaratne I, Menkes DB, Wang X. Chinese non-psychiatric hospital doctors' attitudes toward management of psychological/psychiatric problems. BMC Health Serv Res 2017 Aug 22;17:576. [doi: 10.1186/s12913-017-2521-8]
99. Worthen-Chaudhari L, McGonigal J, Logan K, Bockbrader MA, Yeates KO, Mysiw WJ. Reducing concussion symptoms among teenage youth: Evaluation of a mobile health app. Brain Inj 2017 Jun 30;31(10):1279-1286. [doi: 10.1080/02699052.2017.1332388]
100. Pelletier J, Rowe M, François N, Bordeleau J, Lupien S. No personalization without participation: on the active contribution of psychiatric patients to the development of a mobile application for mental health. BMC Med Inform Decis Mak 2013 Jul 27;13:78 [FREE Full text] [doi: 10.1186/1472-6947-13-78] [Medline: 23890085]
101. Bousquet J, Bewick M, Arnavielhe S, Mathieu-Dupas E, Murray R, Bedbrook A, et al. Work productivity in rhinitis using cell phones: The MASK pilot study. Allergy 2017 Jun 09;72(10):1475-1484. [doi: 10.1111/all.13177]
102. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27;7:MR000042. [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
103. Pombo N, Garcia N, Bousson K, Spinsante S, Chorbev I. Pain Assessment–Can it be Done with a Computerised System? A Systematic Review and Meta-Analysis. Int J Environ Res Pub Health 2016 Apr 13;13(4):415. [doi: 10.3390/ijerph13040415]
104. Udtha M, Nomie K, Yu E, Sanner J. Novel and Emerging Strategies for Longitudinal Data Collection. J Nurs Scholarsh 2014 Dec 09;47(2):152-160. [doi: 10.1111/jnu.12116]
105. Gibson DG, Pereira A, Farrenkopf BA, Labrique AB, Pariyo GW, Hyder AA. Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review. J Med Internet Res 2017 May 05;19(5):e139. [doi: 10.2196/jmir.7428]
106. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
107. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
108. Demirci JR, Bogen DL. An Ecological Momentary Assessment of Primiparous Women's Breastfeeding Behavior and Problems From Birth to 8 Weeks. J Hum Lact 2017 Mar 23;33(2):285-295. [doi: 10.1177/0890334417695206]
109. Faherty LJ, Hantsoo L, Appleby D, Sammel MD, Bennett IM, Wiebe DJ. Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy. J Am Med Inform Assoc 2017 Jul 01;24(4):746-753. [doi: 10.1093/jamia/ocx005] [Medline: 28339686]
110. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile Phone-Delivered Cognitive Behavioral Therapy for Insomnia: A Randomized Waitlist Controlled Trial. J Med Internet Res 2017 Apr 11;19(4):e70. [doi: 10.2196/jmir.6524]
111. Isetta V, Torres M, González K, Ruiz C, Dalmases M, Embid C, et al. A New mHealth application to support treatment of sleep apnoea patients. J Telemed Telecare 2017 Jan;23(1):14-18. [doi: 10.1177/1357633X15621848] [Medline: 26672606]
112. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A Pilot Comparison of a Smartphone App With or Without 2-Way Messaging Among Chronic Pain Patients: Who Benefits From a Pain App? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
113. Noë B, Turner LD, Linden DEJ, Allen SM, Maio GR, Whitaker RM. Timing rather than user traits mediates mood sampling on smartphones. BMC Res Notes 2017 Sep 16;10:481. [doi: 10.1186/s13104-017-2808-1]
114. Silva de Lima AL, Hahn T, Evers LJW, de Vries NM, Cohen E, Afek M, et al. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. PLoS ONE 2017 Dec 20;12(12):e0189161. [doi: 10.1371/journal.pone.0189161]

115. Sundberg K, Wengström Y, Blomberg K, Hälleberg-Nyman M, Frank C, Langius-Eklöf A. Early detection and management of symptoms using an interactive smartphone application (Interaktor) during radiotherapy for prostate cancer. Support Care Cancer 2017 Feb 24;25(7):2195-2204. [doi: 10.1007/s00520-017-3625-8] [Medline: 28236145]
116. Cooray C, Matusevicius M, Wahlgren N, Ahmed N. Mobile Phone–Based Questionnaire for Assessing 3 Months Modified Rankin Score After Acute Stroke. Circ Cardiovasc Qual Outcomes 2015 Oct 29;8(6 Suppl. 3):S125-S130. [doi: 10.1161/circoutcomes.115.002055]
117. Pavliscsak H, Little JR, Poropatich RK, McVeigh FL, Tong J, Tillman JS, et al. Assessment of patient engagement with a mobile application among service members in transition. J Am Med Inform Assoc 2015 Sep 11;23(1):110-118. [doi: 10.1093/jamia/ocv121]
118. Symer MM, Abelson JS, Milsom J, McClure B, Yeo HL. A Mobile Health Application to Track Patients After Gastrointestinal Surgery: Results from a Pilot Study. J Gastrointest Surg 2017 Sep;21(9):1500-1505. [doi: 10.1007/s11605-017-3482-2] [Medline: 28685388]
119. Labhart F, Anderson KG, Kuntsche E. The Spirit Is Willing, But the Flesh is Weak: Why Young People Drink More Than Intended on Weekend Nights-An Event-Level Study. Alcohol Clin Exp Res 2017 Oct 02;41(11):1961-1969. [doi: 10.1111/acer.13490]
120. Olson D, Lamb M, Lopez MR, Colborn K, Paniagua-Avila A, Zacarias A, et al. Performance of a Mobile Phone App-Based Participatory Syndromic Surveillance System for Acute Febrile Illness and Acute Gastroenteritis in Rural Guatemala. J Med Internet Res 2017 Nov 09;19(11):e368. [doi: 10.2196/jmir.8041]
121. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological Momentary Assessment in Behavioral Research: Addressing Technological and Human Participant Challenges. J Med Internet Res 2017 Dec 15;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]
122. Peleg M, Shahar Y, Quaglini S, Broens T, Budasu R, Fung N, et al. Assessment of a personalized and distributed patient guidance system. Int J Med Inform 2017 May;101:108-130. [doi: 10.1016/j.ijmedinf.2017.02.010] [Medline: 28347441]
123. Lepkowski J. Sampling the difficult-to-sample. J Nutr 1991;121(3):416-423. [doi: 10.1093/jn/121.3.416]
124. Rechel B, Mladovsky P, Devillé W. Monitoring migrant health in Europe: A narrative review of data collection practices. Health Policy 2012 Apr;105(1):10-16. [doi: 10.1016/j.healthpol.2012.01.003]
125. General Data Protection Regulation.: intersoft consulting; 2018. URL: https://gdpr-info.eu/
126. WHO. Health of migrants: The way forward, report of a global consultation, Madrid, Spain, 3 - 5 March. Geneva: World Health Organization; 2010.
127. Ingleby D. European research on migration and health: Background paper for the AMAC project. Geneva: World Health Organization; 2009.
128. Zeeb H, Razum O. Epidemiologische Studien in der Migrationsforschung. Bundesgesundheitsbl 2006 Aug 30;49(9):845-852. [doi: 10.1007/s00103-006-0017-5]
129. Reiss K, Makarova N, Spallek J, Zeeb H, Razum O. Identifizierung und Rekrutierung von Menschen mit Migrationshintergrund für epidemiologische Studien in Deutschland. Gesundheitswesen 2012 Aug 29;75(6):e49-e58. [doi: 10.1055/s-0032-1321768]
130. Reiss K, Dragano N, Ellert U, Fricke J, Greiser KH, Keil T, et al. Comparing sampling strategies to recruit migrants for an epidemiological study. Results from a German feasibility study. Eur J Public Health 2014 May 27;24(5):721-726. [doi: 10.1093/eurpub/cku046]
131. Garduño-Diaz SD, Husain W, Ashkanani F, Khokhar S. Meeting challenges related to the dietary assessment of ethnic minority populations. J Hum Nutr Diet 2013 Aug 24;27(4):358-366. [doi: 10.1111/jhn.12153]

Edited by R Kukafka; submitted 04.01.20; peer-reviewed by A Nar, T Miron-Shatz; comments to author 09.09.20; revised version received 06.11.20; accepted 17.12.20; published 22.01.21.

Please cite as:
Fischer F, Kleen S
Possibilities, Problems, and Perspectives of Data Collection by Mobile Apps in Longitudinal Epidemiological Studies: Scoping Review
J Med Internet Res 2021;23(1):e17691
URL: http://www.jmir.org/2021/1/e17691/
doi:10.2196/17691
PMID:33480850

Review

Data Quality and Cost-effectiveness Analyses of Electronic and Paper-Based Interviewer-Administered Public Health Surveys: Systematic Review

Atinkut Alamirrew Zeleke[1,2], PhD; Tolga Naziyok[2], MSc; Fleur Fritz[3], PhD; Lara Christianson[4], MLS; Rainer Röhrig[2,5], MD

[1]Medical Informatics, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
[2]Division of Medical Informatics, Carl von Ossitetzky University Oldenburg, Oldenburg, Germany
[3]Institute of Medical Informatics, University of Münster, Münster, Germany
[4]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[5]Institute for Medical Informatics, Medical Faculty of RWTH University Aachen, Aachen, Germany

Corresponding Author:
Atinkut Alamirrew Zeleke, PhD
Medical Informatics
Institute for Community Medicine
University Medicine Greifswald
Ellernholzstraße 1-2
Greifswald, D-17487
Germany
Phone: 49 3834 86 7548
Email: atinkut.zeleke@uni-greifswald.de

Abstract

Background: A population-level survey (PLS) is an essential and standard method used in public health research that supports the quantification of sociodemographic events, public health policy development, and intervention designs. Data collection mechanisms in PLS seem to be a significant determinant in avoiding mistakes. Using electronic devices such as smartphones and tablet computers improves the quality and cost-effectiveness of public health surveys. However, there is a lack of systematic evidence to show the potential impact of electronic data collection tools on data quality and cost reduction in interviewer-administered surveys compared with the standard paper-based data collection system.

Objective: This systematic review aims to evaluate the impact of the interviewer-administered electronic data collection methods on data quality and cost reduction in PLS compared with traditional methods.

Methods: We conducted a systematic search of MEDLINE, CINAHL, PsycINFO, the Web of Science, EconLit, Cochrane CENTRAL, and CDSR to identify relevant studies from 2008 to 2018. We included randomized and nonrandomized studies that examined data quality and cost reduction outcomes, as well as usability, user experience, and usage parameters. In total, 2 independent authors screened the title and abstract, and extracted data from selected papers. A third author mediated any disagreements. The review authors used EndNote for deduplication and Rayyan for screening.

Results: Our search produced 3817 papers. After deduplication, we screened 2533 papers, and 14 fulfilled the inclusion criteria. None of the studies were randomized controlled trials; most had a quasi-experimental design, for example, comparative experimental evaluation studies nested on other ongoing cross-sectional surveys. A total of 4 comparative evaluations, 2 pre-post intervention comparative evaluations, 2 retrospective comparative evaluations, and 4 one-arm noncomparative studies were included. Meta-analysis was not possible because of the heterogeneity in study designs, types, study settings, and level of outcome measurements. Individual paper synthesis showed that electronic data collection systems provided good quality data and delivered faster compared with paper-based data collection systems. Only 2 studies linked cost and data quality outcomes to describe the cost-effectiveness of electronic data collection systems. Field data collectors reported that an electronic data collection system was a feasible, acceptable, and preferable tool for their work. Onsite data error prevention, fast data submission, and easy-to-handle devices were the comparative advantages offered by electronic data collection systems. Challenges during implementation included technical difficulties, accidental data loss, device theft, security concerns, power surges, and internet connection problems.

Conclusions: Although evidence exists of the comparative advantages of electronic data collection compared with paper-based methods, the included studies were not methodologically rigorous enough to combine. More rigorous studies are needed to compare paper and electronic data collection systems in public health surveys considering data quality, work efficiency, and cost reduction.

International Registered Report Identifier (IRRID): RR2-10.2196/10678

(J Med Internet Res 2021;23(1):e21382) doi:10.2196/21382

KEYWORDS

electronic data collection; demographic and health survey; tablet computer; smartphone; mobile phone

Introduction

Until well-established civil and vital statistics systems are in place in low- and middle-income countries (LMIC), monitoring sociodemographic events using data on vital societal statistics will remain dependent on alternative data sources. Public health surveys—such as censuses, demographic and health surveys (DHS), and health and demographic surveillance—serve as a data lifeline for these countries [1,2]. Mortality and morbidity indicators, service utilization, and population-level program impact evaluations are usually calculated from household-level data. Further analysis of these population-level epidemiologic indicators is helpful in identifying the determinants of mortality and morbidity. Data collection and management is the first step in the process of evidence generation from household surveys, in which data quality errors could be introduced or prevented. Avoiding errors at this stage is the first-line choice to avoid inherited errors in further data management processes [1,3,4].

The current classical data collection and management processes in LMIC are heavily dependent on paper-based manual methods [4,5]. Paper-based data collection requires extensive human and material resources, especially for large-scale surveys. It also incurs high printing and data entry costs and requires extra data quality assurance steps during and after data collection. Moreover, it takes a long time for an error-free data set to be ready for analysis [6,7]. The intrinsic mode of paper-based data collection affects the data quality, timeliness, and cost of survey implementation, among other factors [8-10].

The rapid development of the global telecommunications infrastructure provides an opportunity for mobile and wireless technologies (mobile health [mHealth]) to support health services and research. Harnessing this technology's potential, particularly in LMIC where the disease burden is highest, is becoming a popular strategy led by relevant activities in World Health Organization member countries [11]. There are diverse mHealth solutions broadly categorized as a tool to support communication between health service institutions and individuals. These include health call centers; reminders to attend appointments; providing access to information and education for health care professionals, for example, access to electronic health care databases and clinical decision support systems; and supporting health monitoring and surveillance (eg, data collection and reporting in health surveys, surveillance, and patient monitoring) [12].

The implementation of tablet- or smartphone-based data collection tools is becoming increasingly popular in public health surveys to mitigate challenges encountered in paper-based data collection [13,14]. Compared with face-to-face interviews, a self-administering mode of electronic data collection tools could potentially increase the response rate among stigmatized groups. These tools have been tested in the contexts of drug abuse [15] and sexual health or HIV [16-18] in public health. The findings conclude that respondents prefer electronic data collection tools as a solution for reporting sensitive information.

Considering data collection in clinical trials, electronic clinical report forms (eCRF) show a potential advantage over paper-based clinical case report forms (CRF) [19-22]. Studies have identified the relative advantages of electronic data capturing tools in terms of data quality, timeliness, and implementation cost.

A handful of experience reports are available on the use of electronic data collection methods in health and demographic surveillance systems (HDSS) in the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH) network. The HDSS site in Malawi used an OpenHDS data system as a means of GPS data collection [23]. One surveillance site in Kenya also reported the adoption of technological innovation using OpenHDS to manage a large-scale malaria survey in western Kenya. The findings asserted that electronic data collection (EDC) enabled the collection of demographic and malaria data quickly and effectively. Moreover, the possibility of real-time data quality controls using the system led to an efficient workflow and subsequent cost savings [24]. The Kombewa HDSS in Kenya also collected data electronically using PDAs and computer notebooks [25,26]. Since 2010, the Magu HDSS site in Tanzania has used EDC to enable enumerators to record household information directly in the PDA [27]. The Dabat HDSS site in northwest Ethiopia also reported the use of PDAs as a means of data collection [6,28]. Most HDSS and DHS still use a paper-based data collection system, and those sites with EDC implementation experience have rarely published their experience or the comparative impact of EDC and paper-based data collections. Despite the individual implementation experiences that suggest that EDC tools can improve data quality and work efficiency and reduce overall survey costs, systematic reviews of the available evidence are limited. The focus of the available systematic reviews is primarily on the mixed potential of mHealth, not specifically on the impact of mobile devices on improving the data collection and management processes in surveys [13,29,30]. Therefore, the impacts of EDC tools in surveys need to be separately analyzed and reported. The available Cochrane systematic review on the impact of data quality parameters focuses on self-administered EDC tools and excludes interviewer-administered methods [14]. In the case of face-to-face interviews, the data collection process involves interaction between the questionnaire, respondent, and interviewer. The difference in the mode of questionnaire administration can have serious effects on data quality [9]. Moreover, conducting face-to-face surveys has more organizational costs involved than self-administered surveys.

Therefore, a systematic review that considers interviewer-administered data collection may complement this evidence. We found no systematic review that analyzed the data quality and cost-effectiveness of electronic and paper-based interview-administered public health surveys. The objective of this systematic review is to synthesize the evidence on the effect of using EDC systems on data quality and cost reduction in public health surveys, with a focus on studies reporting comparative impacts of paper-based data collection and EDC.

Methods

We registered a detailed protocol with PROSPERO, an international database of prospectively registered systematic reviews, with the registration number CRD42018092259. PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines were used to report our systematic review [31,32]. The protocol of this study has been published [33].

Inclusion Criteria

We assessed studies that investigated the effect of EDC methods on improving the data quality and cost-effectiveness in public health surveys or surveillance, compared with traditional paper-based data collection methods. We included all mobile apps with technologies that directly support the data collection process by enabling data collectors and interviewers to collect and send data as well as enabling supervisors and data managers to monitor the data collection process. The study participants included in our review are defined as data collection tool users who use a method of data collection.

Studies with the following characteristics were included:

- The study compared either data quality or cost-effectiveness or both as primary outcomes and reported these in the paper.
- The intervention consisted of mobile information and communication technology devices along with mobile apps, which include PDAs, cellphones, smartphones, and tablet computers—devices used specifically for data collection and reporting processes during surveys.
- The control and intervention groups were compared in face-to-face interview-administered surveys conducted at the household level.
- Demographic surveillance sites were based on clinical settings and not mandated for standard clinical trials (eg, CRF vs eCRF).
- The paper was published between January 2008 and December 2018.

Search Information Source and Search Strategies

Studies were identified through systematic searching in the following electronic databases: MEDLINE via Ovid, CINAHL via EBSCO, PsycINFO via Ovid, EconLit via EBSCO, the Social Science Citation Index, the Science Citation Index via the Web of Science and CENTRAL, and the Cochrane Library (Table 1). In addition, the reference lists of all the included citations were screened. We also searched clinical trial registries for unpublished and in-progress studies.

Table 1. Subject term translations for individual databases

MEDLINE and Cochrane	PsycINFO	CINAHL
Mobile applications	Not available	Mobile applications
Computers, handheld	Mobile devices, computer peripheral devices	Computers, handheld
Electronic health records	Not available	Electronic health records
Cell phone	Cellular phones	Cell phone
Surveys and questionnaires	Surveys; questionnaires	Data collection methods
Interviews as topic	Interviews	Included in data collection methods
Costs and cost analysis	Costs and cost analysis	Costs and cost analysis
Data accuracy	Not available	Not available

The search strategy reported in the protocol was refined and updated in collaboration with a research librarian. This strategy considered 3 categories: the technology or intervention used (eg, mobile device, mobile phone, mHealth, or EDC), area of application (eg, data collection, demographic and health survey, or large-scale survey), and the outcome of interest (eg, data quality, missing data, and cost-effectiveness). We linked synonyms and controlled vocabulary with Boolean operators *OR* and the categories with the operator *AND*. Textbox 1 presents the search strategy in MEDLINE via Ovid. Appropriate modifications to control for vocabulary and syntax were made to the search strategy for each database (Textbox 1). Additional search strategies for PsycINFO, CINAHL, Web of Science, and Cochrane databases are presented in the supplementary file (Multimedia Appendix 1). All searches were conducted in January 2019.

Textbox 1. Search strategy in MEDLINE via Ovid.

MEDLINE(R) and Epub Ahead of Print, via Ovid

1. (((tablet or handheld or hand held or electronic) adj2 (device* or computer*)) or ((electronic or digital) adj2 (form? or data capture* or survey* or case report form? or data collection?)) or Open Data Kit or ODK or EDC or eCRF or eHealth or mHealth or digital health or Android or tablet? or PDA? or personal digital assistant? or app? or (mobile adj2 (technolog* or application? or app?)) or ((mobile or cell* or smart) adj2 phone*) or smartphone* or cellphone*).ti,ab.

2. exp “mobile applications”/

3. exp ”computers, handheld”/

4. exp “electronic health records”/

5. exp “cell phone”/

6. or/2-5

7. 1 or 6

8. (field work or fieldwork or HDSS or CAPI or computer assisted personal interviewing or questionnaire* or survey* or interview* or (population adj2 surveillance) or DHS or EDC or (data adj2 (gather* or captur*)) or (health and demographic surveillance system?)).ti,ab.

9. exp “surveys and questionnaires”/

10. exp “interviews as topic”/

11. or/9-10

12. 8 or 11

13. ((cost? adj2 (analy?s or comparison* or saving? or measure? or effectiv* or reduction? or reduce? or reduction or reducing or decrease? or decreasing)) or (cost benefit adj2 (analy?s or comparison* or measure?)) or (cost utility adj2 (analy?s or comparison* or measure?)) or economic evaluation? or quality control? or (data adj2 (quality or accuracy or accurate* or error? or error rate? or incomplete* or complete* or inaccurate* or inaccuracy or valid*))).ti,ab.

14. exp “costs and cost analysis”/

15. exp “data accuracy”/

16. or/14-15

17. 13 or 16

18. 7 and 12 and 17

19. Limit 18 to yr=“2008 -Current”

Study Selection

We imported all citations from all databases to EndNote for deduplication management and further screening. Although we planned to use the Covidence web-based screening tool to manage the title and abstract screening process, we finally chose the Rayyan QCRI (Qatar Computing Research Institute) screening tool because it is freely available and provides sufficient screening functionalities. Two authors (AZ, MPH in Health Informatics, and TN, MSc in Informatics) independently screened the titles, abstracts, and full text, based on the inclusion criteria. Disagreements and uncertainty on the screening results were first resolved through discussion among the reviewers, followed by consultation with the third (FF, Postdoc in Medical Informatics) and fifth authors (RR, Professor in Medical Informatics). We used a pretested and standardized (through calibration exercise) Microsoft Excel sheet for data extraction based on the inclusion criteria and the objectives of the review.

Data Management and Extraction Process

Two reviewers extracted the following information from the papers:

- Bibliographic information (authors, titles, journals, and year of publication).
- Characteristics of the intervention (eg, hardware, software, and networking).
- Study methods (design, setting, participants, and sample size).
- Assessed outcomes (data quality, cost-effectiveness, and others).
- Quantitative or qualitative summary of the main findings, including descriptive frequencies and statistical tests.

The full description of the data extraction items can be accessed in the published protocol [33].

Risk of Bias or Quality Assessment

Randomized controlled trials are suitable for evaluating whether drugs are effective; however, for interventions that involve health care delivery modes, it may not be appropriate or possible to conduct a randomized controlled trial. We aimed to assess the quality of the data in the included studies using parameters such as random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other biases. However, the included studies were neither randomized controlled trials nor nonrandomized trials with clinical outcomes; they were mainly prospective comparative experimental studies, cross-sectional studies, or historical secondary data record comparisons. The remaining studies were a one-time feasibility study or experience reports from the implementation or use of an EDC tool in public health practice.

Data Synthesis

There was substantial heterogeneity among the studies concerning the intervention (mobile electronic data capturing tools such as PDA, smartphone, or tablet computer and the app they used), outcome types (error rate and missing or inaccurate data), and level of outcome measurements (sample level, household level, and variable level) of the mHealth interventions and study outcomes.

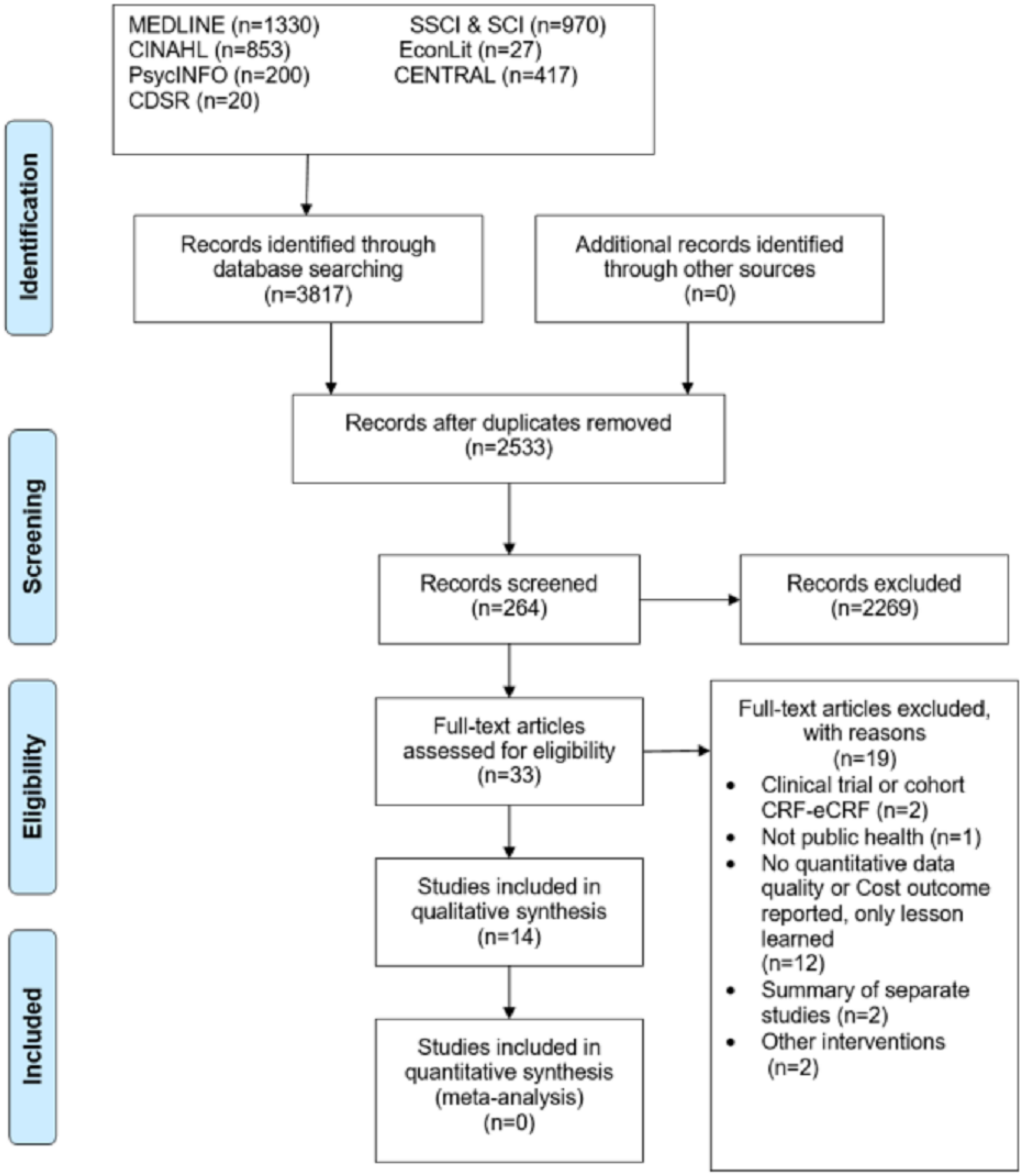
The studies were found to be noncombinable, and combining these studies would not have been methodologically sound. Consequently, we performed a narrative synthesis of the studies.

Results

Study Selection and Characteristics

The search performed in the included databases yielded 3817 results. After deduplication, 2533 results were exported to the Rayyan QCRI screening tool (Figure 1). Of these, 2500 papers were discarded after title and abstract screening, as these papers clearly did not meet our criteria. Of the 33 full-text papers included, only 14 met the amended inclusion criteria. The original protocol was aimed at including comparative studies that addressed paper-based and electronic tools in the same study, conducted household-level data collection in a community field setting, and reported primary outcomes (data quality or cost-effectiveness) of both data collection tools in the same paper. Only 7 studies (that are heterogeneous) fulfilled these criteria. Due to the limited evidence, we extended our inclusion criteria to cover studies that use the tools in demographic surveys or surveillance systems in a clinical or hospital setting. We also included one-sided study design papers that only reported primary outcomes (cost or data quality) and EDC methods, without formal comparison with paper-based data collection methods. This widening of the scope provided an additional 7 papers (3 comparative and 4 noncomparative EDC papers) to bring the total to 14: 10 comparative and 4 noncomparative single-arm studies were included for final full-text extraction (Figure 1).

Figure 1. Screening process based on the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) flowchart template.



Study Characteristics

The final systematic synthesis analysis included 14 studies conducted in 12 LMIC. Of these 14, 10 [7,34-42] comparatively assessed the outcomes using paper-based data collection and EDC in the same analysis. Furthermore, 4 papers [43-46] reported either data quality or cost-related outcomes or both in a study conducted using an EDC tool (Table 2).

Table 2. Study and content characteristics of the included papers

Category	Studies, n	References
Country (n=14)		
Kenya	2	[36,45]
Ethiopia	1	[7]
China	2	[42,43]
Malawi	1	[39]
India	1	[37]
Philippines/Bangladesh	1	[34]
Sudan	1	[35]
Burkina Faso	2	[40,44]
Tanzania	1	[38]
South Africa	1	[41]
Nigeria	1	[46]
Study setting (n=10)		
Household community setting	7	[7,34,35,37-40]
Clinical/hospital setting	3	[36,41,42]
Comparison of paper-based and electronic data collection (n=10)		
Both from the same study	8	[34,35,37-42]
Compared from 2 studies conducted at different times	2	[7,36]
Purpose of the study design (n=7)		
Primarily designed to evaluate paper-based and electronic tools	5	[35,37,39,41,42]
Secondary byproduct of another primary survey	2	[7,36]
Types of outcomes (n=10)		
Only data quality outcomes	1	[35]
Only cost outcomes	2	[34,40]
Both cost and data quality outcomes	7	[7,36-39,41,42]
Level of data quality outcome assessment (n=8)		
Household level	2	[7,38]
Questionnaire level	4	[7,35,41,42]
Variable level	3	[36,37,39]
Type of data quality outcome comparison (n=8)		
Missing	4	[7,36,38,39]
Inaccurate	4	[7,36,38,39]
Mixed (identified as error)	5	[35,37,38,41,42]
Economic evaluation type (n=11)		
Complete input cost	3	[36,38,40]
Partial (differential) cost	8	[7,37,39,41-43,45,46]
Usability/user preference evaluation (n=14)		
Reported after formal evaluation	6	[7,34,37-39,42]
Reported with informal discussion	1	[36]
No user evaluation information	7	[35,40,41,43-46]
Study, intervention, or evaluation year (n=10)		
2008-2012	5	[36,38,41,42,44]

Category	Studies, n	References
2013-2018	5	[34,35,37,43,46]
Publication year (n=14)		
2008-2012	2	[42,46]
2013-2018	12	[7,34-41,43,45,46]

Furthermore, 5 studies [35,37,39,41,42] were primarily intended to evaluate and compare data quality and cost-related outcomes from a prospective study design using paper-based and electronic tools. The remaining 5 papers [7,34,36,38,40] reported the outcomes from previous technology utilization experiences. The reported outcomes were not primarily intended to evaluate the tools; rather, data quality and cost-related outcomes were extracted from surveys at different times.

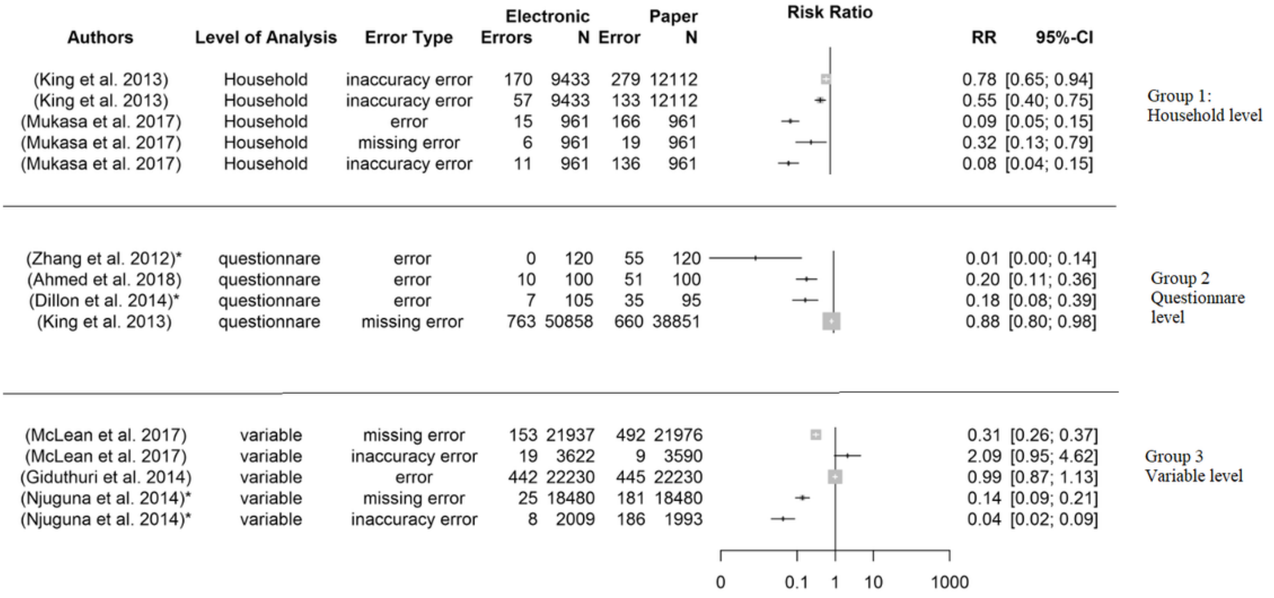
Regarding the settings, 7 studies [7,34,35,37-40] in the comparative group included data from a household-level survey, and 3 studies [36,41,42] conducted surveys in clinical settings or research centers.

Types of Outcomes

Data Quality Outcomes

Data quality outcomes, as defined in the methods section, comprise the frequency of errors (incomplete, missing, or inaccurate items) on 3 levels, based on the reported outcomes at the household, questionnaire, and variable levels. At the household level, the incidence of 1 or more errors among the total number of households included in the surveys and data analysis were reported in 2 papers [7,38]. Similarly, the frequency of 1 or more errors per complete questionnaire, regarded as a questionnaire-level error, was reported in 4 studies [7,35,41,42]. At a variable level, a count of the errors in a complete set of questions or variables in questionnaires measured as a variable error were mentioned in 3 papers (Figure 2) [36,37,39].

Figure 2. Forest plot comparison of heterogeneity characteristics of the data quality outcomes.



The cost of implementing electronic and paper-based data collection processes was estimated by most studies (12 out of 14). The majority of these estimated the partial or differential costs unique to that study and its EDC [7,37,39,41-43,45,46]. Only 3 papers listed the full implementation cost [36,38,40] for both the study and the EDC. Except 2 studies that compared costs per correctly entered data observation or error-free databases [39,41], none attempted to link the data quality outcome measures to the cost inputs.

Effect on Data Quality: Missing Data and Inaccuracy Errors

Errors are reported in terms of missing and inaccurate data in 4 studies [7,36,38,39], while a combination of both errors as a single error indicator is reported in 4 others [35,37,38,42] (Table 3).

Table 3. Extracted data quality outcomes.

Study and country	Type of comparison, study design, and setting[a]	Study population and sample unit	Intervention: device or app	Methods of paper-based and electronic tool administration	Outcome measurement characteristics[b]	Result	Biases (selection, information, or confounding)
Ahmed et al [35], Sudan	B: Exploratory pilot study nested and experimented in cross-sectional household surveys	• Sample unit: respondents • Sample size: 100 for each (PPDC[c] and EDC[d])	EDC: Smartphone ODK[e] app	Daily sample randomization for EDC or PPDC. 1 respondent simultaneously interviewed by 2 interviewers with different tools (PPDC and EDC)	• A: Data quality errors, questions with no answers, or wrong use of the skip pattern • B: After data entry • C: Errors measured at the questionnaire level • D: Used similar questionnaires • (BOLD[f] survey core questionnaire)	• PPDC: 51 errors from 100 questionnaires • EDC: 10 errors from 100 electronic forms • 83% errors in PPDC	Selection and information
King et al [7], Ethiopia	A: Pre- and postdesign in full EDC implementation evaluation in cross-sectional household surveys	Sample units: • Households • PPDC: 9433 • EDC: 12,112 • Individuals enumerated in PPDC: 38,851 and in EDC: 50,858 • Individuals examined in PPDC: 33,800 and in EDC: 38,652	EDC: For example, tablet computer or self-developed Android app	Paper and electronic surveys at different places and times, 1 data collector at a time	• A: Data quality, percentage of individuals enumerated with at least 1 blank field in the census record • Percentage of households with incorrect unique identifying number (inaccuracy at the household level) • GPS with blank entries (missing at the household level) • The proportion of total time • B: After data entry (raw data sets) • C: At the questionnaire and household levels • D: Used the same questionnaire (Trachoma Impact Evaluation Survey questionnaire)	• PPDC: Missing rate at the questionnaire level is 1.7% • EDC: Missing 1.5% at P=.01 • PPDC: Missing rate in GPS data 0.6% (N=9263) • EDC: Missing rate in GPS 1.1% (N=12,064; P<.01; person-days) • PPDC: Proportion of total time 790 person-days • EDC: Proportion of total time 511 person-days	Selection, information, and confounding

Study and country	Type of comparison, study design, and setting[a]	Study population and sample unit	Intervention: device or app	Methods of paper-based and electronic tool administration	Outcome measurement characteristics[b]	Result	Biases (selection, information, or confounding)
McLean et al [39], Malawi	A: Prospective evaluation nested and experimented in cross-sectional household surveys at HDSS[g] sites	- PPDC: 426 respondents - EDC: 558 respondents - Time: 3 weeks	EDC: Tablet computer and smartphone, ODK app	Independent parallel EDC or PPDC, 1 data collector at a time	- A: Missing data defined as not asked (blank; discounting *not applicable* blank questions) or as blank and entered as unknown combined, Internal validity: defined as a field with an impossible or inconsistent value and time for submission - B: Not clear - C: Variable-level error count - D: Used different questionnaires	- PPDC: Missing 492 (2.2%) of 21,976 fields - EDC: Missing 153 (0.7%) of 21,937 fields (RR[h] 3.2, 95% CI 2.7-3.8) - PPDC: Internal inconsistencies in 19 (0.5%) of 3590 fields - EDC: Internal inconsistencies in 9 (0.2%) of 3622 fields (RR 0.5, 95% CI 0.2-1.1) - Time from interview to data availability on the database - PPDC: mean 3.4 days (range 3.0-3.7) - EDC: mean 2.1 days (range 2.0-2.3)	Selection and information
Giduthuri et al [37], India	A: Prospective experimental comparative study designed for households selected randomly	- 98 interviews in EDC and PPDC	EDC: Tablet computer/ODK app	A respondent simultaneously interviewed by 2 interviewers with different tools (PPDC and EDC)	- A: Discrepancies with reference to device-attributable errors - Paper entries incorrect or missing and tablet entries missing because paper interviewer (lead) did not follow the skip logic - Tablet entries incorrect or missing, paper entries missing because tablet interviewer's (lead) skip logic was considered an error of EDC - B: During data entry - C: Variable-level error count - D: Used the same questionnaire	- PPDC: Mean number of paper-attributable errors 4.68 (445/22,230, 2.01%) - EDC: Mean number of tablet-attributable errors 4.65 (442/22,230, 1.99%)	Selection, information, and confounding

Study and country	Type of comparison, study design, and setting[a]	Study population and sample unit	Intervention: device or app	Methods of paper-based and electronic tool administration	Outcome measurement characteristics[b]	Result	Biases (selection, information, or confounding)
Mukasa et al [38], Tanzania	A: Retrospective record review of household survey data	• 961 households for EDC and PPDC	EDC or PDA BlackBerry customized HRS[i]	Repeated survey, PPDC first, followed by EDC, 1 data collector at a time	• A: Error category from the database: accuracy, logic, and consistency; range; and completeness and missing values • B: After data entry from database • C: Error identified at the household level • D: Used the same questionnaire	• Households with errors • PPDC: 166 (17%) • EDC: 15 (2%) • PPDC more likely with accuracy errors (79%; 95% CI 72%-86%); EDC: (58%; 95% CI 29%-87%) • Errors in EDC more likely to be related to completeness (32%; 95% CI 12%-56%) than in PPDC (11%; 95% CI 7%-17%)	Selection, information, and confounding
Zhang et al [42], China	B: Prospective comparison study conducted in a clinic	• 60 mothers for each tool (EDC and PPDC)	EDC: smartphones, leased software	A respondent simultaneously interviewed by 2 interviewers with same tools (PPDC or EDC). Random assignment of respondents to one of the tools	• A: Data quality errors; missing confirmation of default option, survey date, missed questions, 2 options circled, wrong options chosen, date, ID; database completion (data entry, checking, and data cleaning) • B: Before and after data entry • C: Questionnaire-level error count • D: Used the same questionnaire	• Households with errors • ◦ PPDC: 55/120 ◦ EDC: 0 • Questionnaire pairs with no recording variations • ◦ EDC: 15/60 ◦ PPDC: 20/60 • EDC: 134 of 186 variables (72.0%) did not have any recording variation • PPDC: 126 of 184 variables (68.5%) did not have any recording variation • In data entry: 65.0% (78/120) of the PPDC records did not match database completion time • PPDC: 16 hours; EDC: 30 min	Selection, information, and confounding

Study and country	Type of comparison, study design, and setting[a]	Study population and sample unit	Intervention: device or app	Methods of paper-based and electronic tool administration	Outcome measurement characteristics[b]	Result	Biases (selection, information, or confounding)

Njuguna et al [36], Kenya	A: Pre- and postimplementation evaluation in hospital-based surveillance data collection	- EDC: 880 questionnaires (May 2011-June 2012)
		- PPDC: 880 questionnaires (January 2010-June 2011)
			EDC: smartphones, FAST[j] app	Paper and electronic surveys at the same place but different times, 1 data collector at a time	- A: Incomplete records or variables in selected questions in a questionnaire; questions requiring responses with or without programmed checks on the smartphone version; percentage of erroneous and inconsistent responses in the questionnaires
					- B: Error count from original responses in the paper
					- C: Variable-level error count
					- D: Used the same questionnaire
						- Missing error:
						- - EDC: 25/18,480
						  - PPDC: 181/18,299
						- Inaccuracy error:
						- - EDC: 8/2009
						  - PPDC: 186/1807
							Selection, information, and confounding

Dillon et al [14], South Africa	B: Prospective data collection in case control study in a hospital setting	- EDC: 105 respondents - PPDC: 95 respondents	EDC: tablet computer, self-developed app	One data collector interviewed all respondents either with paper or with electronic tools in a random order	- A: Data collection errors defined as nonsensical or impossible inputs, missing data, or inputs inconsistent or incompatible with previous responses during the interview; minor errors classified as differences of 1 year or less in date calculations. Major errors classified as all other error types - B: Not clear - C: Variable-level error count - D: Used similar questionnaires	- Overall number of errors per 100 questions - EDC: 0.17 errors; PPDC: 0.73 errors - P<.001 - Interview duration (EDC: 5.4 min; PPDC: 5.6 min)	Selection, information, and confounding

---

[a] A: full implementation; B: pilot testing.

[b] A: type of outcome; B: stage of error assessment for paper questionnaire; C: error measurement level; D: questionnaire name/s similarities.

[c] PPDC: pen and paper data collection.

[d] EDC: electronic data collection.

[e] ODK: Open Data Kit.

[f] BOLD: Burden of Obstructive Lung Disease.

[g] HDSS: health and demographic surveillance systems.

[h] RR: risk ratio.

[i] HRS: Household Registration System.

[j] FAST: Field Adapted Survey Toolkit.

An exploratory pilot study nested and experimented in a cross-sectional household survey in Sudan compared data quality errors, questions with no answers, or incorrect use of the skip pattern in 100 convenience samples. A pair of data collectors simultaneously interviewed each respondent—1 with pen and paper and 1 with electronic tools—and recorded the data separately. In the paper-based data collection, 51 of the 100 questions had one or more errors, compared with 10 errors in the electronically submitted forms [35].

A study in India by Giduthuri et al [37] also compared error rates between paper-based and electronically collected data from a comparative prospective experimental study. The data collectors were randomly assigned to use either pen and paper or EDC tools while interviewing each respondent simultaneously. Audio-recorded data during the survey were used as a reference to compare discrepancies and device-attributable errors. According to the reference, paper errors indicate when a paper entry was incorrect and when a tablet entry was missing because the paper-based tool interviewer (lead) did not follow the skip logic. Furthermore, tablet entries were incorrect or missing and paper entries were missing because of the electronic tool interviewer's (lead) skip logic was considered an error of the electronic data collection tool. The mean number of paper-attributable errors was 4.68 (445/22,230, 2.01%), while the mean number of tablet-attributable errors was 4.65 (442/22,230, 1.99%); thus, no significant differences were observed [37].

A study in China compared smartphone and paper-based data collection in an infant feeding practice survey conducted in rural clinics. Purposive sampling techniques were used to select 120 mothers, 60 per survey tool group. Two data collectors with the same tool (paper or electronic) interviewed 60 mothers in random order, yielding 120 records for each tool. For the paper-based questionnaire, 55 of 120 questionnaires had 1 or more errors or missing data. The most frequent error was a missing confirmation of the default option, which was observed 156 times in 49 questionnaires. No missing error was reported for the EDC tool group [42]. The mean duration of an interview was 10.22 (SD 2.17) min for the smartphone method and 10.83 (SD 2.94) min for the pen and paper method. Moreover, database completion took 16 hours (including data entry, checking, and data cleaning) for pen and paper data collection (PPDC), while it took half an hour for EDC [42].

A prospective evaluation experiment conducted a nested, ongoing, cross-sectional household survey at HDSS sites in Malawi. In 3 weeks, 426 interviews with PPDC and 558 interviews with EDC were conducted. Data collectors independently interviewed different households in a *1 data collector in 1 home* mode. Missing data were defined as not asked (blank; discounting *not applicable* blank questions) or as blank and entered as unknown combined. Internal validity was defined as a field with an impossible or inconsistent value and time for submission. In paper questionnaires, missing data were reported in 492 (2.2%) of 21,976 fields, compared with 153 (0.7%) of 21,937 fields in electronic forms (risk ratio [RR] 3.2, 95% CI 2.7-3.8). Internal inconsistencies were found in 19 (0.5%) of 3590 fields collected by PPDC compared to 9 (0.2%) of 3622 fields for EDC (RR 0.5, 95% CI 0.2-1.1) [39]. Moreover, the mean data availability duration in databases was 3.4 days (95% CI 3.0-3.7) in PPDC compared with 2.1 days (95% CI 2.0-2.3) for EDC. The mean number of interviews per day was similar for both groups at 10.7 (95% CI 8.7-12.6) for PPDC and 11.8 (95% CI 8.1-15.5) for EDC [39].

A tablet computer-based data collection system was implemented in a large-scale study of trachoma impact assessment surveys in Ethiopia [7]. Data quality outcomes were compared with a similar paper-based survey conducted 7 months earlier in a different part of the country. The sampling units were households (PPDC: 9433 vs EDC: 12,112), and the study enumerated 38,851 individuals in the PPDC survey and 50,858 in the EDC survey. Individuals enumerated with at least 1 blank field in a single respondent response were defined as missing data (1.7% for PPDC vs 1.5% for EDC; P=.01). Missing data at the household level for GPS with blank entries was also reported (EDC: 1.1% vs PPDC: 0.6%; P<.01). Inaccuracy errors were defined only in a percentage of households with an incorrect unique identifying number (PPDC: 2.3% vs EDC: 1.8%). Data entry and analysis were done in less than 1 day for EDC, while it took 14 days for data entry and an additional 5 days for double data entry discrepancy checks for PPDC [7].

Apart from the above studies, interview time or mean data availability duration in the databases were also reported in some papers. A study in Kenya reported faster data upload to a central database in EDC and a median duration for data upload of 7 days (range 1-13 days) after data collection for EDC and 21 days (range 4-56 days) for PPDC (P<.01) [36]. A combined report from Indonesia and the Philippines showed that the median time between data collection and data entry for PPDC surveys was approximately 3 months, compared with 2 days for EDC [34]. Time analysis from a large-scale survey in Tanzania reported that the median duration of an enumeration session per household was 9.4 min (90% central range 6.4-12.2) for paper surveys and 8.3 min (6.1-12.0) for electronic surveys (P=.01) [38].

Effect on Cost-effectiveness

Most of the studies reported cost analysis for the expenses incurred to conduct surveys using paper-based and electronic tools. However, the included studies varied significantly in the types and level of cost analysis reported in their groups. Most of the recommendations from the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement [47] were not included. In Table 4, we provide basic information on the study and country, analytic method and model, participants per intervention, time horizon, discount rate, currency, included cost inputs, cost ranges, outcomes, consequence, and conclusion information.

Table 4. Extracted cost information (based on Consolidated Health Economic Evaluation Reporting Standards evaluation template)

Study and country	Analytic method or model	Interventions studied or population per group (1=PPDC[a]; 2=EDC[b])	Time horizon, discount rate, currency (base year)	Included cost inputs and assumptions (1=PPDC; 2=EDC)	Data quality outcome link with input cost	Cost range of intervention (1=PPDC; 2=EDC)	Conclusions and remark
King et al [7], Ethiopia	Input cost analysis	- 9433 - 12,112	- NR[c] - NR - US $ (2011)	- Printing and data entry costs - Tablet and accessories Single-use cost	Not linked	1. US $13,883 2. US $10,320	- Costs of the equipment in EDC approximately the same as with data entry cost, recurrent use of EDC may save costs. Use of person day for comparison of personnel costs
McLean et al [39], Malawi	Input cost analysis Differential cost	- 426 - 558	- 1 year - NR - British Pound (2016)	- Printing and entry cost and paper archival - EDC development and configuration, device cost, data officer cost	Not linked	1. £18,895 2. £11,427	- In total, the estimated costs for the stages unique to the paper-based process is 65% higher per annum than the unique costs for the EDC system
Giduthuri et al [37], India	Differential input cost analysis	- 98 interviews for both, and extrapolated for larger samples 1000	- NR - NR - US $ (2013)	- Printing and data entry expenses - Cost of tablet computers and server charges	Not linked	For 96 interviews, the cost is 1. US $2598 2. US $2648	- The initial investment in tablet-based interviews was higher compared to paper, while the recurring costs per interview were lower with the use of tablets - EDC is less expensive for larger surveys
Mukasa et al [38], Tanzania	Cost and cost-effectiveness	- 1000 households for both	- NR (used deflator values) - NR (used the number of households) - US $ (2008)	- All fixed costs and running costs - To estimate the costs for 2015, the formula for expenditure in 2016 = Expenditure in 2008 × Deflator 2016, deflator 2008 - Deflator values of 209.5 for 2008 and 233.6 for 2016	Error rate	For 1000 households, error-free data set: 1. US $1161 2. US $9380 Crude data set: 1. US $993 2. US $891	- For error-free data sets, surveys using electronic tools, compared with paper-based tools, were less expensive by 28% for recurrent and 19% for total costs

Study and country	Analytic method or model	Interventions studied or population per group (1=PPDC[a]; 2=EDC[b])	Time horizon, discount rate, currency (base year)	Included cost inputs and assumptions (1=PPDC; 2=EDC)	Data quality outcome link with input cost	Cost range of intervention (1=PPDC; 2=EDC)	Conclusions and remark

Zhang et al [42], China	Input cost analysis	• 60
		• 60
		• Extrapolated to
		  • 1200
		  • 1600
			• NR
			• NR (used projected samples)
			• US $ (2012)
				• Both EDC and PPDC: Items for preparation, training, fieldwork and data collection, and logistics
				• Only printing and transporting the questionnaire, stationery, and data entry
				• Only renting the smartphone and the software
				• Assumptions for the extrapolation are not known
					Not linked	Sample size: 60 each
						1. US $1500
						2. US $2700
						Sample size: 1200
						1. US $41,570
						2. Cost per sample, US $23.77
						Sample size: 1600
						1. US $28,520
						2. Cost per sample, US $25.98
							• The mean costs per questionnaire were higher for the smartphone questionnaire
							• No linked analysis for data quality and cost

Njuguna et al [36], Kenya	Input cost analysis	• Both 880 • Extrapolated for the first year's establishing and second year's operating cost	• NR • 2 years • US $ (2011, 2012)	• First- and second-year costs of starting up and operating based on payroll information	Not linked	First year: 1. US $15,999 2. US $17,500 Second year: 1. US $54,001 2. US $50,200	• For establishment cost, EDC 9.4% more than PPDC; in 2 years, EDC costs reduced 7% compared to PPDC • No linked analysis for data quality and cost
Dillon et al [41], South Africa	Input and economic analysis	• 95 • 105	• British pound (2012)	• Salary per correctly entered question • Technology costs and tablet computer and additional overhead costs (storage space for paper hard copies; office space for data entry clerk, and for EDC, hardware maintenance and upkeep) • Formulas presented by Walther et al [19]. • Assumed minimum staffing: • 1 field worker, 1 data entry clerk, and 1 data supervisor • 1 field worker and 1 data manager (EDC) • Calculation based on 46 questions in a questionnaire, 5 interviews per day, 22 working days per month, 110 interviews per month, and 5060 questions per month	Error rate	Salary cost per month: 1. £1000 2. £915 Equipment cost: 1. £420 2. £1036	• EDC salary cost per correctly entered question is 0.5 times that of PPDC • Initial technology costs for the EDC is 2.47 times that of PPDC • Cost per question • • EDC: £0.18 • PPDC: £0.20 • EDC saved £101.20 per month • EDC cost recoup time 6 months

Study and country	Analytic method or model	Interventions studied or population per group (1=PPDC[a]; 2=EDC[b])	Time horizon, discount rate, currency (base year)	Included cost inputs and assumptions (1=PPDC; 2=EDC)	Data quality outcome link with input cost	Cost range of intervention (1=PPDC; 2=EDC)	Conclusions and remark
Flexman et al [34], Bangladesh and Philippines	Input data analysis	• 5398 • 516	• 2013 verbal autopsies' data collection in demographic and surveillance sites	• Differential cost estimation formula excluding similar cost for EDC and PPDC, for example, data collector cost (see the full paper mentioned on page 4) • The cost for data entry for a single enterer is the one-time cost of a computer plus the monthly salary multiplied by the number of months required	Not linked	• Printing cost per paper questionnaire US $0.246 (Bangladesh) and US $0.774 (Philippines) • Cost of a single electronic tablet US $393.78 (Bangladesh) and US $365.76 (Philippines) • Cost of a computer for data entry US $984.49 (Bangladesh) and US $1000.00 (Philippines) • Monthly salary of data enterer US $384.62 (Bangladesh) and US $326.00 (Philippines) • Average number of samples per person per month 107.8 (Bangladesh) and 145.7 (Philippines)	• For small-scale surveys, the upfront costs of purchasing electronic tablets was the primary cost, and it had a higher total cost. For large-scale surveys, the costs associated with data entry exceeded the cost of the tablets, so • EDC provides a cheaper method of data collection • Historical cost data from 2 countries. For projects that require fewer than 150 tablets and collect over 10,000 surveys, the upfront cost of the tablets will likely be substantially less than the cost of data entry
Lietz et al [40], Burkina Faso	Comparative input cost analysis	• 10,000 households for both	• HDSS[d] and HMS[e] 2010 and inflated to 2014 values • CDA[f] actual pretesting expenditure in 2014	• Fixed costs: personnel (team lead), office, and housekeeping • Variable costs: data collectors, supervisors, consumables, transportation, and training (full text in Textbox 1) • Fixed and variable costs Financial costs of the standalone (HDSS and HMS) and integrated (CDA) survey approaches were estimated from the perspective of the implementing agency	Not assessed	Cost per household visit 1. HMS + HDSS (PPDC): US $251,641 2. CDA (EDC): US $206,937 Cost per (10,000 HDSS) 1. EDC: US $21 2. PPDC: US $25	• Cost analysis estimated that the CDA survey would reduce the annual costs of survey implementation by about US $45,000 • No link with the data quality outcome

[a] PPDC: pen and paper data collection.

[b] EDC: electronic data collection.

[c] NR: not reported

[d] HDSS: health and demographic surveillance systems.

[e] HMS: Household Morbidity Survey.

[f] CDA: Comprehensive Disease Assessment.

Studies from Ethiopia [7], Malawi [39], India [37], and Bangladesh and Philippines [34] reported a differential input cost unique to paper-based or electronic tools. For paper-based data collection, these were printing and data entry costs, and for EDC systems, the cost of electronic devices' hardware, software, and accessories. The general assumption was that all other costs, such as personnel costs, were the same for both tools. Another cost assumption was that of the cost of small-scale, short-duration surveys. Such kinds of small costs were extrapolated to large-scale surveys with no clear information about the model or the cost assumptions followed to reach the large-scale costs. The studies concluded that EDC was expensive for small-scale surveys, as the initial investment in hardware and software outweighs the paper-based printing and data entry costs. However, large-scale surveys showed a significant decrease in cost for EDC surveys; for example, the paper-based survey cost was up to 65% higher per annum than the unique costs for the EDC system [39]. None of these studies linked the input cost with data quality errors.

Detailed cost information and the link between cost and data quality were reported in a retrospective data analysis in Tanzania. However, the base year for the cost was 2008, and deflator values of 209.5 for 2008 and 233.6 for 2016 were used to report the costs. For 1000 households, the cost per error-free data set was US $11,610 for PPDC and US $9380 for EDC. For error-free data sets, surveys using electronic tools—compared with paper-based tools—were 28% less expensive in recurrent costs and 19% less expensive in total costs [38].

A study in South Africa also reported cost inputs linked to data quality outcomes [41]. The formula presented by Walther et al [19] was used with minimum staffing: for PPDC, 1 field worker, 1 data entry clerk, and 1 data supervisor, and for EDC, 1 field worker and 1 data manager. In addition, 46 questions in a questionnaire, 5 interviews per day, 22 working days per month, 110 interviews per month, and 5060 questions per month were planned. The EDC salary cost per correctly entered question was 0.5 times that of PPDC. Overall, the cost per question was £0.18 for EDC and £0.20 for PPDC. The equipment cost for PPDC was £420, compared with £1036 for EDC. EDC saved £101.20 per month, and the EDC cost recoup time was reported as 6 months.

Lingani et al [40] in Burkina Faso and Njuguna et al [36] in Kenya reported a detailed financial cost comparison for the establishment of PPDC and EDC at HDSS and a hospital-based surveillance system, respectively. The Kenyan report indicated that during establishment, the cost of EDC was 9.4% higher than that of PPDC. However, after 2 years, EDC costs decreased by 7%, compared with PPDC (see Table 4 for detailed cost information).

Technology Characteristics, User Preference, and Acceptance

The mobile devices used for data collection included PDAs, smartphones, tablet computers, and notebooks (Table 5). The included studies (6/12, 50%) also reported the use of open source Android apps called open data kit apps to customize the software according to their needs. Microsoft Windows and BlackBerry operating systems were installed on mobile devices.

Data transfer from mobile devices to the central server was conducted using one of the following methods:

- Direct transfer from the data collection site to the server using a mobile data network [34-36] and secure virtual private network [45].
- Direct transfer using Wi-Fi connection only [37,39].
- Transfer using secure digital memory card [7,38].
- Transfer using USB cable [41].

Data transfer to a server located in a foreign country using a mobile network was not considered appropriate in some studies due to data ownership or security concerns [7,43].

Table 5. Extracted intervention or technology characteristics

Study and country	Analytic method or model	Interventions studied or population per group (1=PPDC[a]; 2=EDC[b])	Time horizon, discount rate, currency (base year)	Included cost inputs and assumptions (1=PPDC; 2=EDC)	Data quality outcome link with input cost
Njuguna et al [36], Ethiopia	• HTC smartphone • Software: FAST[a] kit: Microsoft • Purchased	Direct to server with mobile network	• Programmed checks and restrictions • Error message notification for inaccurate data entry	• Limited or poor internet network • Occasional server communication breakdowns • Delayed data submission	Data saved in the smartphone's memory and later uploaded on to the server from convenient places
Ahmed et al [35], Sudan	• Samsung smartphone • Customized ODK[b] • Open source	Mobile internet network	• ODK functionalities	• Limited or poor internet network in certain areas • Unexpected software closure or freeze • Accessibility of the aggregate website in Sudan was challenging • Using Arabic language in the ODK was challenging for form development and data retrieval	Used ODK offline and submitted the data after restarting the smartphone
King et al [7], Ethiopia	• Samsung tablet computer • Swift Insights Mobile 1.1 • Barcode scanner 4.3.1 Open source developed	SD[c] card with password-protected downloading to the supervisor's laptop	• ODK functionalities+ User-defined survey preferences, generation of unique record identification (Textbox 1)	• Limited or poor internet network in certain areas	Data stored first on the supervisor laptop and then uploaded to the local server to maintain the sovereignty and security of the data set
Flexman et al [34], Bangladesh and Philippines	• Samsung tablets • ODK • Customized in ODK	Direct to server with mobile network	• ODK functionalities (no explicit description)	• Sufficiently strong internet network	Purchase memory cards for the tablets to back up data locally
Giduthuri et al [37], India	• Samsung Note external recorder • Enhanced ODK development	Encrypted and uploaded over a Wi-Fi connection to a central server after returning to the office	Not described	• Required highly trained interviewers • Data loss by the interviewer due to accidentally pressing the delete button	Not described
Musaka et al [38], Tanzania	• PDA BlackBerry • HRS[d] software • BlackBerry OS[e] • Customized in HRS software	Micro SD card	• Skip function	• Device stopped functioning during interview or submission	Not described
McLean et al [39], Malawi	• Toshiba tablets and Samsung smartphones • ODK • Contractual payment for form development	Secured wireless network (not relayed on phone network)	• ODK functionalities	• The existing wireless network was sufficiently strong to upload the data • Adequate battery life for a day • Lack of internal staff for form development • Required outsourcing cost • 11 of 92 tablets broke in 4 years and had to be replaced • Only 1 functioning tablet and 1 smartphone went missing over the 4-year period	A dedicated, secured device charging area

Study and country	Analytic method or model	Interventions studied or population per group (1=PPDC[a]; 2=EDC[b])	Time horizon, discount rate, currency (base year)	Included cost inputs and assumptions (1=PPDC; 2=EDC)	Data quality outcome link with input cost
Dillon et al [41], South Africa	• Tablet, PC, and mobile phone • Self-developed a C#-based program with XML	Data transfer through USB connections, avoiding the need for a constant internet connection	• Facilitate data checks and early detection and correction of faulty procedures and data management	• A well-run staff training program • Staff was already technically trained • Lack of reliable internet connection	Data transfer through USB
Jing et al [43], China	• Smartphone • ODK and importing and extraction software • ODK Briefcase v 1.4 Production	Stored safely on the smartphone and uploaded to a computer secured by a password. Compiled data sent to the country coordinator	• ODK functionalities	• Interviewers do not have direct access to submit the data to the server in Washington, forcing them to store the data on the smartphones for a short period • Loss of data due to smartphone damage • Uploading data was a challenge for older smartphone users	Building a server in China that can be easily accessed would facilitate improved data security and immediate assignment of cause of death on smartphones at the time of interview
Byass et al [44], Burkina Faso	• PDA • Pendragon 4 software programmed	Data copied for PDAs' memory cards	• GPS, time stamp	• Use several available mobile-charging arrangements in vehicles or solar panels • Temporary system clock changes in a few PDAs • Some incomplete GPS data strings • 48/151 PDA encountered technical errors	Saving data from the PDAs' internal volatile memory to nonvolatile memory cards. Protective plastic cover for the safety of PDA was important
Maduga et al [46], Nigeria	• Mobile phone • ODK • Android • Locally customized	Trained to upload completed forms onto a secure server, with back-end access provided to only the research team lead	• ODK functionalities	• Limited network connection • Fluctuation of power and internet connection	The extra 2 phones served as backup in the event of malfunction or challenges with the global system for mobile communication mobile and data networks. Multiple SIM cards were provided in an attempt to mitigate the problem

[a] FAST: Field Adapted Survey Toolkit.

[b] ODK: Open Data Kit.

[c] SD: Secure Digital.

[d] HRS: Household Registration System.

[e] OS: operating system.

Technical Challenges in Electronic Data Collection

Limited or poor internet connectivity, occasional server communication interruptions, language challenges, the need for highly trained data collectors, device stack or freezing, device loss and breakdowns, and limited battery or power sources are among the technological challenges faced in the implementation of EDC systems (Table 5). The solutions include offline storage and transfer as soon as the data collectors obtain reliable internet connection (store and forward methods), transferring data using USB cables or secure digital cards, purchasing backup mobile devices or batteries, and using paper questionnaires at times of device malfunction.

Preference, Acceptability, and Feelings

A total of 6 papers [7,34,37-39,42] reported that user preferences, acceptance, and opinions were assessed using formal evaluation methods, such as individual or focus groups or qualitative interviews. Detailed comparative advantages and disadvantages of paper and electronics tools are reported in thematic-based tables in [7,38,39].

The use of smartphones to collect data was faster, easier to follow, and more convenient, as the data collectors did not have to carry cumbersome paper questionnaires and less space was needed to store their data collection tools. Additional functionalities—such as automatic retrieval of respondents and other members of the household or GPS functionalities—are also reported as an advantage.

The risk of data loss with paper-based surveys was perceived as being less than that with EDC, as paper questionnaires are tangible and enable immediate review, identification, and correction of mistakes. Paper surveys were also perceived to be easier for manipulating, adding, or changing data—for instance, including a household member absent during the survey who was later encountered by the survey team. The automated skip function was advantageous and time saving. The enumerators did not have to read the questions on every visit to the same household.

Enumerators described that the devices felt exciting, interesting, and prestigious, and they were skilled professionals in the eyes of the community. Some fieldworkers felt that EDC interviews took longer, and occasionally, devices froze during an interview.

Some studies that were excluded from our review also offer important insights about preference, acceptability, and local experiences [48-51].

Discussion

This systematic review synthesized the available comparative evidence on paper-based and electronic data collection tools and the potential effect of using these tools on data quality, implementation cost, and user preferences in interview-administered public health surveys. The systematic review included studies from 2008 to 2018 that were identified through multiple online electronic database searches. We identified more than 3500 papers and screened more than 2500 titles and abstracts to include 14 full-text papers based on our inclusion criteria. We extracted and synthesized available evidence regarding data quality, cost-effectiveness, timeliness, and user preferences. No paper has reported a study design with a classical randomized control approach. Randomization was reported to indicate either respondent allocation to paper-based or electronic tools or data collectors' exposure to one of the data collection tools. Meta-analysis was not possible due to the heterogeneous nature of study designs, measurements and outcome types, and study settings. Instead, a narrative synthesis based on predefined data quality, cost and related outcomes' acceptability, and preferences was conducted.

We employed a rigorous systematic review process to formulate the research questions, prepare individual database-tailored search strategies, execute searches in multiple databases, and independently screen and extract evidence from thousands of papers.

However, the results were inadequate for meta-analysis. The final included studies were heterogeneous and could not be combined to generate better estimates. Low-quantity and low-quality phenomena are becoming evident in most recent systematic reviews assessing mHealth or eHealth outcomes [13,29,30].

This scarcity might result from the following reasons: lack of primary studies with a rigorous study design, insufficient search strategy or review process, or unnecessary narrowing of a study focus. The most commonly reported reason is a lack of sufficient, well-planned, rigorous studies. The lack of evidence might be due to reluctance to evaluate the system after implementation and due to publication bias because of unsuccessful or disappointing findings [52]. Apart from a few studies [35,37,39,41,42], those included in our systematic review were not primarily designed to evaluate the impact of EDC compared with PPDC. Instead, the studies were a byproduct of survey experience from a comparative analysis using their secondary data. Information from such implementation practices can provide insights or lessons for readers, but the comparative outcomes might be influenced by unplanned and uncontrolled confounding variables [7,36]. Elimination of observed and unobserved factors that might otherwise plausibly explain the difference in outcomes in the study design can increase confidence in the assertion that the estimated impact constitutes the real impact of the tools. The studies included in this review suffered from multiple biases during sample size estimation, selection (purposive vs random selection), and data quality outcome measurement level (before or after data entry). We recommend that future research focus specifically on the mode of data collection measurement and on quantifying the impacts with sound research designs.

Generating a full economic evaluation of the evidence facilitates a comparison between interventions in terms of their costs and intended outcomes and can be used to inform decisionmakers or funders of the available choices among alternatives upon cost justification [53,54]. In our systematic review, we attempted to extract the available cost information using the CHEERS checklist [47]; however, most of the expected items in the checklist were not reported. A majority of the studies lacked a detailed description of unit costs, data sources, and cost calculations. Moreover, most used a time horizon of 1 year and failed to assess long-term costs and data quality effects. The rationale for the choice of the time horizon was also not explicitly stated.

Despite these limitations, the available cost data could provide clues regarding the existing cost parameters for paper- and electronic data collection systems. Two studies managed to link the cost of implementing EDC and PPDC tools to data quality. Such cost-effectiveness analyses should be encouraged in future studies. There is no clear answer or guideline to shape the type and level of rigorous studies in health information technology evaluation [55,56]. Details of evidence-based health informatics history, current practices, and future recommendations are discussed elsewhere [52]. Further debate and consensus among academics and researchers in biomedical informatics should continue to determine how and when health information technology evaluation is rigorous and produces good quality data [57].

The studies identified in this review were conducted in various countries and in the context of different health care systems. Generalizing and applying results from different contexts is difficult because of variations in clinical practice, costs, and their analysis. However, what was consistent across all studies was a lack of reporting on the feasibility of adopting these technologies based on economic and organizational factors.

It was surprising to see limited publications from global survey implementation organizations for the DHS and INDEPTH network groups, as those projects have many years of multinational implementation experience [4,58]; however, apart from experience reports, comparative evaluation studies in those areas are rare. Further evaluation research in these projects might produce evidence of data quality, cost, timeliness, and the success and failure factors for multinational projects.

There are positive perceptions regarding the acceptability, usability, and preference of EDC over PPDC among data collectors. This positivity is because technology enables data collectors to focus on their work, get immediate feedback regarding their mistakes, correct their errors while in the field, and leave few data quality issues to revisit. It is not known whether this excitement is a short-term effect immediately after the technology introduction or a stable long-term view based on longer exposure. A short period of technophilic or technophobic attitudes might lead to inaccurate overall impression of the tool, as accurate impression can only develop with longer exposure to the technology [59]. The generalizability and applicability of the results, given the different types of devices with different technical specifications and the rapid pace at which technology advances, need critical evaluation. The generalizability of the findings of this systematic review is also challenged by the studies themselves, considering the variations in the characteristics of data collectors, level of outcome measurements, settings of the survey, and the psychometric properties of the survey questionnaires.

Conclusions

This systematic review showed that, despite consistent claims of a positive impact of technology on data quality and cost-effectiveness, the available evidence is small in quantity and low in quality. Purposefully designed comparative studies assessing the impact of data quality and cost-effectiveness are needed for implementation in organizations and by decision makers.

Despite the heterogeneity and low quality of the included studies, their qualitative synthesis showed the superiority of EDC systems over paper-based systems for data quality, process efficiencies, and cost.

Comparative evaluation studies sourced from international survey-implementing organizations where their routine data collection mode is EDC can provide a better platform for impact evaluation research in large-scale surveys.

Acknowledgments

The authors would like to thank the Deutscher Akademischer Austauschdienst for funding the research stay of the principal investigator.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional search strategies.
[DOCX File , 17 KB - jmir_v23i1e21382_app1.docx]

References

1. Bocquier P, Sankoh O, Byass P. Are health and demographic surveillance system estimates sufficiently generalisable? Glob Health Action 2017;10(1):1356621 [FREE Full text] [doi: 10.1080/16549716.2017.1356621] [Medline: 28820344]
2. Timaeus I, Harpham T, Price M, Gilson L. Health surveys in developing countries: the objectives and design of an international programme. Soc Sci Med 1988;27(4):359-368. [doi: 10.1016/0277-9536(88)90270-5] [Medline: 3175718]
3. Gerritsen A, Bocquier P, White M, Mbacké C, Alam N, Beguy D, et al. Health and demographic surveillance systems: contributing to an understanding of the dynamics in migration and health. Glob Health Action 2013 Jul 11;6(1):21496 [FREE Full text] [doi: 10.3402/gha.v6i0.21496] [Medline: 23849188]
4. Survey Organization Manual for Demographic and Health Surveys. ICF International. 2012. URL: https://dhsprogram.com/pubs/pdf/DHSM10/DHS6_Survey_Org_Manual_7Dec2012_DHSM10.pdf [accessed 2018-12-05]
5. Ethiopia Demographic and Health Survey. Central Statistical Agency [Ethiopia] and ICF International. 2012. URL: https://dhsprogram.com/countries/Country-Main.cfm?ctry_id=65 [accessed 2018-10-02]
6. Okwaraji YB, Cousens S, Berhane Y, Mulholland K, Edmond K. Effect of geographical access to health facilities on child mortality in rural Ethiopia: a community based cross sectional study. PLoS One 2012;7(3):e33564 [FREE Full text] [doi: 10.1371/journal.pone.0033564] [Medline: 22428070]
7. King JD, Buolamwini J, Cromwell EA, Panfel A, Teferi T, Zerihun M, et al. A novel electronic data collection system for large-scale surveys of neglected tropical diseases. PLoS One 2013;8(9):e74570 [FREE Full text] [doi: 10.1371/journal.pone.0074570] [Medline: 24066147]
8. Toninelli D, Pinter R, Pablo DP. Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies. London, UK: Ubiquity Press; 2015.

9. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291. [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]
10. Groves R. Survey Errors and Survey Costs. New Jersey: John Wiley & Sons; 2004.
11. World Health Organization. Global Diffusion of Ehealth: Making Universal Health Coverage Achievable. Report of the Third Global Survey on Ehealth. Geneva, Switzerland: World Health Organization; 2016.
12. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug 6;1(2):160-171. [doi: 10.9745/ghsp-d-13-00031]
13. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
14. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison C, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27(7):MR000042. [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
15. Perlis TE, Des Jarlais DC, Friedman SR, Arasteh K, Turner CF. Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. Addiction 2004 Jul;99(7):885-896. [doi: 10.1111/j.1360-0443.2004.00740.x] [Medline: 15200584]
16. Adebajo S, Obianwu O, Eluwa G, Vu L, Oginni A, Tun W, et al. Comparison of audio computer assisted self-interview and face-to-face interview methods in eliciting HIV-related risks among men who have sex with men and men who inject drugs in Nigeria. PLoS One 2014;9(1):e81981 [FREE Full text] [doi: 10.1371/journal.pone.0081981] [Medline: 24416134]
17. Beauclair R, Meng F, Deprez N, Temmerman M, Welte A, Hens N, et al. Evaluating audio computer assisted self-interviews in urban South African communities: evidence for good suitability and reduced social desirability bias of a cross-sectional survey on sexual behaviour. BMC Med Res Methodol 2013 Jan 31;13:11 [FREE Full text] [doi: 10.1186/1471-2288-13-11] [Medline: 23368888]
18. Richens J, Copas A, Sadiq ST, Kingori P, McCarthy O, Jones V, et al. A randomised controlled trial of computer-assisted interviewing in sexual health clinics. Sex Transm Infect 2010 Aug;86(4):310-314. [doi: 10.1136/sti.2010.043422] [Medline: 20551234]
19. Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. PLoS One 2011;6(9):e25348 [FREE Full text] [doi: 10.1371/journal.pone.0025348] [Medline: 21966505]
20. Le Jeannic A, Quelen C, Alberti C, Durand-Zaleski I, CompaRec Investigators. Comparison of two data collection processes in clinical studies: electronic and paper case report forms. BMC Med Res Methodol 2014 Jan 17;14:7 [FREE Full text] [doi: 10.1186/1471-2288-14-7] [Medline: 24438227]
21. Pavlović I, Kern T, Miklavcic D. Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study. Contemp Clin Trials 2009 Jul;30(4):300-316. [doi: 10.1016/j.cct.2009.03.008] [Medline: 19345286]
22. Rorie DA, Flynn RW, Grieve K, Doney A, Mackenzie I, MacDonald TM, et al. Electronic case report forms and electronic data capture within clinical trials and pharmacoepidemiology. Br J Clin Pharmacol 2017 Sep;83(9):1880-1895 [FREE Full text] [doi: 10.1111/bcp.13285] [Medline: 28276585]
23. di Pasquale A, McCann RS, Maire N. Assessing the population coverage of a health demographic surveillance system using satellite imagery and crowd-sourcing. PLoS One 2017;12(8):e0183661 [FREE Full text] [doi: 10.1371/journal.pone.0183661] [Medline: 28859109]
24. Homan T, Di Pasquale A, Kiche I, Onoka K, Hiscox A, Mweresa C, et al. Innovative tools and OpenHDS for health and demographic surveillance on Rusinga Island, Kenya. BMC Res Notes 2015 Sep 1;8:397 [FREE Full text] [doi: 10.1186/s13104-015-1373-8] [Medline: 26323664]
25. Kaneko S, K'opiyo J, Kiche I, Wanyua S, Goto K, Tanaka J, et al. Health and demographic surveillance system in the western and coastal areas of Kenya: an infrastructure for epidemiologic studies in Africa. J Epidemiol 2012;22(3):276-285 [FREE Full text] [doi: 10.2188/jea.je20110078] [Medline: 22374366]
26. Sifuna P, Oyugi M, Ogutu B, Andagalu B, Otieno A, Owira V, et al. Int J Epidemiol 2014 Aug;43(4):1097-1104 [FREE Full text] [doi: 10.1093/ije/dyu139] [Medline: 25009309]
27. Kishamawe C, Isingo R, Mtenga B, Zaba B, Todd J, Clark B, et al. Int J Epidemiol 2015 Dec;44(6):1851-1861 [FREE Full text] [doi: 10.1093/ije/dyv188] [Medline: 26403815]
28. Okwaraji YB, Webb EL, Edmond KM. Barriers in physical access to maternal health services in rural Ethiopia. BMC Health Serv Res 2015 Nov 4;15:493 [FREE Full text] [doi: 10.1186/s12913-015-1161-0] [Medline: 26537884]
29. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
30. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]

31. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
32. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
33. Zeleke AA, Naziyok T, Fritz F, Röhrig R. Data quality and cost-effectiveness analyses of electronic and paper-based interviewer-administered public health surveys: protocol for a systematic review. JMIR Res Protoc 2019 Jan 30;8(1):e10678 [FREE Full text] [doi: 10.2196/10678] [Medline: 30698530]
34. Flaxman AD, Stewart A, Joseph JC, Alam N, Alam SS, Chowdhury H, et al. Collecting verbal autopsies: improving and streamlining data collection processes using electronic tablets. Popul Health Metr 2018 Feb 1;16(1):3 [FREE Full text] [doi: 10.1186/s12963-018-0161-9] [Medline: 29391038]
35. Ahmed R, Robinson R, Elsony A, Thomson R, Squire SB, Malmborg R, et al. A comparison of smartphone and paper data-collection tools in the Burden of Obstructive Lung Disease (BOLD) study in Gezira state, Sudan. PLoS One 2018;13(3):e0193917 [FREE Full text] [doi: 10.1371/journal.pone.0193917] [Medline: 29518132]
36. Njuguna HN, Caselton DL, Arunga GO, Emukule GO, Kinyanjui DK, Kalani RM, et al. A comparison of smartphones to paper-based questionnaires for routine influenza sentinel surveillance, Kenya, 2011-2012. BMC Med Inform Decis Mak 2014 Dec 24;14:107 [FREE Full text] [doi: 10.1186/s12911-014-0107-5] [Medline: 25539745]
37. Giduthuri JG, Maire N, Joseph S, Kudale A, Schaetti C, Sundaram N, et al. Developing and validating a tablet version of an illness explanatory model interview for a public health survey in Pune, India. PLoS One 2014;9(9):e107374 [FREE Full text] [doi: 10.1371/journal.pone.0107374] [Medline: 25233212]
38. Mukasa O, Mushi HP, Maire N, Ross A, de Savigny D. Do surveys with paper and electronic devices differ in quality and cost? Experience from the Rufiji Health and demographic surveillance system in Tanzania. Glob Health Action 2017;10(1):1387984 [FREE Full text] [doi: 10.1080/16549716.2017.1387984] [Medline: 29157182]
39. McLean E, Dube A, Saul J, Branson K, Luhanga M, Mwiba O, et al. Implementing electronic data capture at a well-established health and demographic surveillance site in rural northern Malawi. Glob Health Action 2017;10(1):1367162 [FREE Full text] [doi: 10.1080/16549716.2017.1367162] [Medline: 28922071]
40. Lietz H, Lingani M, Sié A, Sauerborn R, Souares A, Tozan Y. Measuring population health: costs of alternative survey approaches in the Nouna Health and Demographic Surveillance System in rural Burkina Faso. Glob Health Action 2015;8:28330 [FREE Full text] [doi: 10.3402/gha.v8.28330] [Medline: 26257048]
41. Dillon DG, Pirie F, Rice S, Pomilla C, Sandhu MS, Motala AA, African Partnership for Chronic Disease Research (APCDR). Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa. J Clin Epidemiol 2014 Dec;67(12):1358-1363 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.06.012] [Medline: 25135245]
42. Zhang S, Wu Q, van Velthoven MH, Chen L, Car J, Rudan I, et al. Smartphone versus pen-and-paper data collection of infant feeding practices in rural China. J Med Internet Res 2012 Sep 18;14(5):e119 [FREE Full text] [doi: 10.2196/jmir.2183] [Medline: 22989894]
43. Zhang J, Joshi R, Sun J, Rosenthal SR, Tong M, Li C, et al. A feasibility study on using smartphones to conduct short-version verbal autopsies in rural China. Popul Health Metr 2016;14:31 [FREE Full text] [doi: 10.1186/s12963-016-0100-6] [Medline: 27547126]
44. Byass P, Hounton S, Ouédraogo M, Somé H, Diallo I, Fottrell E, et al. Direct data capture using hand-held computers in rural Burkina Faso: experiences, benefits and lessons learnt. Trop Med Int Health 2008 Jul;13(Suppl 1):25-30 [FREE Full text] [doi: 10.1111/j.1365-3156.2008.02084.x] [Medline: 18578809]
45. Ojwang JK, Lee VC, Waruru A, Ssempijja V, Ng ang a JG, Wakhutu BE, et al. Using information and communications technology in a national population-based survey. J Acquired Immun Def Syndrome 2014;66:S123-S129. [doi: 10.1097/qai.0000000000000116]
46. Maduka O, Akpan G, Maleghemi S. Using android and open data kit technology in data management for research in resource-limited settings in the niger delta region of nigeria: cross-sectional household survey. JMIR Mhealth Uhealth 2017 Nov 30;5(11):e171 [FREE Full text] [doi: 10.2196/mhealth.7827] [Medline: 29191798]
47. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, CHEERS Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Br Med J 2013 Mar 25;346:f1049 [FREE Full text] [doi: 10.1136/bmj.f1049] [Medline: 23529982]
48. Seebregts CJ, Zwarenstein M, Mathews C, Fairall L, Flisher AJ, Seebregts C, et al. Handheld computers for survey and trial data collection in resource-poor settings: development and evaluation of PDACT, a Palm Pilot interviewing system. Int J Med Inform 2009 Nov;78(11):721-731. [doi: 10.1016/j.ijmedinf.2008.10.006] [Medline: 19157967]
49. Paudel D, Ahmed M, Pradhan A, Lal Dangol R. Successful use of tablet personal computers and wireless technologies for the 2011 Nepal Demographic and Health Survey. Glob Health Sci Pract 2013 Jul 11;1(2):277-284. [doi: 10.9745/ghsp-d-12-00056]

50. Knipe DW, Pearson M, Borgstrøm R, Pieris R, Weerasinghe M, Priyadarshana C, et al. Challenges and opportunities of a paperless baseline survey in Sri Lanka. BMC Res Notes 2014 Jul 15;7(1):-. [doi: 10.1186/1756-0500-7-452]
51. Sime H, Deribe K, Assefa A, Newport MJ, Enquselassie F, Gebretsadik A, et al. Integrated mapping of lymphatic filariasis and podoconiosis: lessons learnt from Ethiopia. Parasit Vectors 2014 Aug 27;7:397 [FREE Full text] [doi: 10.1186/1756-3305-7-397] [Medline: 25164687]
52. Rigby M, Ammenwerth E. The need for evidence in health informatics. Stud Health Technol Inform 2016;222:3-13. [Medline: 27198087]
53. Schweitzer J, Synowiec C. The economics of eHealth and mHealth. J Health Commun 2012 May 02;17 Suppl 1(sup1):73-81. [doi: 10.1080/10810730.2011.649158] [Medline: 22548602]
54. LeFevre AE, Shillcutt SD, Broomhead S, Labrique AB, Jones T. Defining a staged-based process for economic and financial evaluations of mHealth programs. Cost Eff Resour Alloc 2017;15:5 [FREE Full text] [doi: 10.1186/s12962-017-0067-6] [Medline: 28428734]
55. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mhealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 9;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
56. van Poucke S, Thomeer M, Heath J, Vukicevic M. Are randomized controlled trials the (g)old standard? From clinical intelligence to prescriptive analytics. J Med Internet Res 2016 Jul 6;18(7):e185 [FREE Full text] [doi: 10.2196/jmir.5549] [Medline: 27383622]
57. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010 Nov 2;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
58. Ethiopia Demographic Health Survey 2016. Central Statistical Agency (CSA) [Ethiopia] and ICF. 2016. URL: https://dhsprogram.com/pubs/pdf/FR328/FR328.pdf [accessed 2018-08-12]
59. Edith L. The Effect of Computer-assisted Interviewing on Data Quality: a Review of the Evidence. Utrecht University Repository. 2008. URL: https://dspace.library.uu.nl/handle/1874/44502 [accessed 2020-12-30]

Abbreviations

CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CRF: clinical case report form
DHS: demographic and health survey
eCRF: electronic clinical report form
HDSS: health and demographic surveillance system
INDEPTH: International Network for the Demographic Evaluation of Populations and Their Health
mHealth: mobile health
RR: risk ratio

Edited by G Eysenbach; submitted 12.06.20; peer-reviewed by R Lundin, M Rauws; comments to author 19.07.20; revised version received 03.08.20; accepted 28.10.20; published 22.01.21.

Please cite as:
Zeleke AA, Naziyok T, Fritz F, Christianson L, Röhrig R
Data Quality and Cost-effectiveness Analyses of Electronic and Paper-Based Interviewer-Administered Public Health Surveys: Systematic Review
J Med Internet Res 2021;23(1):e21382
URL: http://www.jmir.org/2021/1/e21382/
doi:10.2196/21382
PMID:33480859

Review

Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for Processing Free Text in Health Care: Systematic Scoping Review

Christophe Gaudet-Blavignac[1,2], BSc, MSc; Vasiliki Foufi[1,2], PhD; Mina Bjelogrlic[1,2], PhD; Christian Lovis[1,2], MPH, MD, FACMI

[1]Division of Medical Information Sciences, Geneva University Hospitals, Geneva, Switzerland

[2]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland

Corresponding Author:
Christophe Gaudet-Blavignac, BSc, MSc
Division of Medical Information Sciences
Geneva University Hospitals
Rue Gabrielle-Perret-Gentil 4
Geneva, 1205
Switzerland
Phone: 41 22 372 62 01
Email: christophe.gaudet-blavignac@hcuge.ch

Abstract

Background: Interoperability and secondary use of data is a challenge in health care. Specifically, the reuse of clinical free text remains an unresolved problem. The Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) has become the universal language of health care and presents characteristics of a natural language. Its use to represent clinical free text could constitute a solution to improve interoperability.

Objective: Although the use of SNOMED and SNOMED CT has already been reviewed, its specific use in processing and representing unstructured data such as clinical free text has not. This review aims to better understand SNOMED CT's use for representing free text in medicine.

Methods: A scoping review was performed on the topic by searching MEDLINE, Embase, and Web of Science for publications featuring free-text processing and SNOMED CT. A recursive reference review was conducted to broaden the scope of research. The review covered the type of processed data, the targeted language, the goal of the terminology binding, the method used and, when appropriate, the specific software used.

Results: In total, 76 publications were selected for an extensive study. The language targeted by publications was 91% (n=69) English. The most frequent types of documents for which the terminology was used are complementary exam reports (n=18, 24%) and narrative notes (n=16, 21%). Mapping to SNOMED CT was the final goal of the research in 21% (n=16) of publications and a part of the final goal in 33% (n=25). The main objectives of mapping are information extraction (n=44, 39%), feature in a classification task (n=26, 23%), and data normalization (n=23, 20%). The method used was rule-based in 70% (n=53) of publications, hybrid in 11% (n=8), and machine learning in 5% (n=4). In total, 12 different software packages were used to map text to SNOMED CT concepts, the most frequent being Medtex, Mayo Clinic Vocabulary Server, and Medical Text Extraction Reasoning and Mapping System. Full terminology was used in 64% (n=49) of publications, whereas only a subset was used in 30% (n=23) of publications. Postcoordination was proposed in 17% (n=13) of publications, and only 5% (n=4) of publications specifically mentioned the use of the compositional grammar.

Conclusions: SNOMED CT has been largely used to represent free-text data, most frequently with rule-based approaches, in English. However, currently, there is no easy solution for mapping free text to this terminology and to perform automatic postcoordination. Most solutions conceive SNOMED CT as a simple terminology rather than as a compositional bag of ontologies. Since 2012, the number of publications on this subject per year has decreased. However, the need for formal semantic representation of free text in health care is high, and automatic encoding into a compositional ontology could be a solution.

(J Med Internet Res 2021;23(1):e24594) doi:10.2196/24594

KEYWORDS

SNOMED CT; natural language processing; scoping review; terminology

Introduction

Background

The ability to meaningfully exchange and process data is of utmost importance in health care, whether it is inside a hospital setting either among different health structures or among health systems in different countries [1-3]. The use of a common terminology is a way to improve both interoperability and the secondary use of data [4].

The Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) was created in 1999 by the fusion of 2 important health care terminologies—SNOMED reference terminology (SNOMED RT) and Clinical Terms Version 3. It was first released in 2002. SNOMED CT is currently considered as the most comprehensive, multilingual, clinical health care terminology in the world, with more than 350,000 concepts and a million relationships [5-7]. It is maintained and published by SNOMED International, a nonprofit organization comprising 39 member countries [8]. In the last 18 years, SNOMED CT has grown in size and coverage and has been included as a standard vocabulary in the meaningful use program [9]. This is an important step for any electronic health record willing to attain interoperability.

With 3 components, namely concepts, descriptions, and relationships, SNOMED CT can be observed as both a complex ontology and a graph containing vertices and labeled edges. This structure allows interesting features such as compositional grammar, expression constraint queries, or postcoordination. It is therefore possible to create postcoordinated concepts that represent new meanings not present in the terminology. These postcoordinated concepts can then be queried and processed with the rest of the terminology [5,10,11].

These characteristics, similar to those of a natural language, make SNOMED CT a candidate for representing clinical free text in a semantically rich, machine-readable manner. Although encoding free text into SNOMED CT can be done manually, it is costly and not scalable for large data sets. Therefore, it is often accomplished by natural language processing (NLP). NLP is an active research branch in the biomedical field and has been broadly applied in the scientific literature and clinical text for diverse tasks [12-14]. However, NLP applications on clinical documents are less frequent. Among the reasons explaining this disparity are the limited access to corpora of clinical documents and the lack of publicly available annotated corpora [15]. These barriers are even more important for languages other than English.

Objectives

SNOMED CT has already been the subject of many studies and evaluations of its coverage, ability to represent complex concepts, or usability in a clinical setting [16-19]. Its usage has already been a subject of reviews; however, those publications are older than 10 years [13,20] or focus on its general use without focusing on its usage to process and represent unstructured data such as clinical free text [7]. Therefore, this work aims to better understand the use of SNOMED CT for representing free text in medicine via a scoping systematic review. It also aims to decipher the use of this terminology across fields, languages, and countries and how it is used from an analytical point of view, such as terminology source up to exploiting its advanced features, that is, postcoordination and compositional grammar.

Methods

Article Selection Process

An exploratory research performed using text-based queries on MEDLINE and Google Scholar helped in defining the queries, topics, and objectives of this study. This work led to the selection of 3 databases for the review based on previous reviews addressing similar topics [7,20,21]. This choice was made to increase coverage. Purely engineering-related databases, such as the Institute of Electrical and Electronics Engineers Xplore or the Association for Computing Machinery digital library, were not selected because of the technical content of their publications, which was often not related to real clinical settings.

In this work, clinical free text is considered as any text written in a natural language about a patient, which does not come from a finite value set. Free-text fields in structured forms and problem lists have been included to broaden the scope.

The selected databases were PubMed [22], Embase [23], and Web of Science [24]. The final query used was as follows: (“SNOMED-CT” OR “SNOMED CT”) AND (“free-text” OR “free text” OR “narrative”). These keywords were defined during the preliminary research. The bottleneck was the presence of the term “SNOMED CT,” and no other synonyms of narrative or free text were added as they did not change the results. The final query was made on August 9, 2019.

To be selected, an article must meet the following inclusion criteria:

- It should be published in scientific journals or conference proceedings after 2002.
- It should include the usage of SNOMED CT to represent or process clinical free text.

The limitation on the date was set to avoid publications that focused on the previous versions of SNOMED.

Although the selection was voluntarily broad, white papers, editor papers, posters, or conference abstracts were excluded. Articles not available in English were also excluded. The Unified Medical Language System (UMLS) [25] developed by the National Library of Medicine (NLM) combines biomedical terminologies in a single resource. Since the release of the UMLS-labeled 2004AA [26,27], it contains SNOMED CT. In this work, publications focusing on the usage of UMLS were included only if they specifically mentioned the usage of SNOMED CT.

To be as inclusive as possible on the chosen topic, the references in every publication were also reviewed to include new publications. The recursive reference review was stopped when no additional publications were added to the set. This has been done with the aim of reducing the impact of the query on the final selection of articles. Moreover, 3 review articles about information extraction from clinical free text were included in the selection. Despite not meeting the inclusion criteria, they were considered as a source of reference to other publications meeting the criteria. Obviously, they were not the target of the topic review described below.

Topics Reviewed

The articles were then studied to extract some specific topics in a systematic manner. The first topic reviewed was the type of document used as a free-text source. To better detect which data were used in these publications, we defined the categories described in Textbox 1.

Textbox 1. Categories of documents.

- History and physical examinations: this category includes documents summarizing the situation of a patient admitted in a health care structure, and his or her physical examination such as admission notes
- Clinical summaries: this category includes any document summarizing a care episode such as a discharge summary
- Death certificates
- Problem lists: this category regroups documents listing the problems of a patient admitted in a health care structure
- Autopsy reports
- Incident reports
- Allergy reports
- Complementary exam reports: this category regroups any document related to a complementary exam, including but not limited to radiology, pathology, and genomic reports
- Narrative notes: this category includes progress notes, nurse notes, and clinical notes not further specified
- Various: this category was selected when a publication used more than one type of document according to this classification

The publications were then classified according to the language they targeted in their work. All the selected publications included a part where the free text was mapped to SNOMED CT concepts. This terminology binding step was classified depending on its justification and whether it was the final goal of the research or a step toward another goal. Textbox 2 defines the types of reasons. These reasons have been defined empirically to fully cover the possibilities encountered in publications. For each type, a point was added if it was present in the publication.

The method used for the terminology binding to SNOMED CT was classified as “manual,” “rule-based,” “machine learning,” or “hybrid” for each article. The definitions used for these categories are listed in Textbox 3. When mapping was accomplished using a specific software, it was reviewed.

The general usage of SNOMED CT was reviewed on 2 specific topics: whether the full terminology or a subset of concepts was used and whether more advanced features of SNOMED CT were included in the study.

Textbox 2. Categories classifying the reason for the terminology binding to Systematized Nomenclature of Medicine Clinical Terms.

- Information extraction: Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is used to extract meaningful information from free text. The focus must be aimed at extracting information, not structuring or encoding it. Publications using the terminology binding to extract clinical information from documents that fall under this category
- Data normalization: SNOMED CT is used to encode existing data. This category is different from information extraction because it focuses on adding semantics to the data while keeping it intact. It includes publications where SNOMED CT is used to define a template or to support information entry
- Synonym resource: SNOMED CT includes synonyms for a large number of its concepts. In this category, SNOMED CT is used as a source for synonyms
- Quality evaluation: SNOMED CT is used to evaluate the quality of care or documentation
- Coverage evaluation: The focus is aimed at evaluating the coverage of SNOMED CT for a specific task by mapping it to free text
- Similarity evaluation: SNOMED CT is used to evaluate similarity among data. It is usually made by using the relationships present in SNOMED CT to compute the semantic distance between concepts
- Gold standard creation: SNOMED CT is used to create a gold standard data set
- Feature in a classification task: SNOMED CT mapping is used as a feature in a classification task
- Value set creation: SNOMED CT is used to define a specific value set
- Mapping to other terminologies: SNOMED CT is used as a bridge to other terminologies

Textbox 3. Definition of the categories used to classify the mapping method.

- Manual: the mapping is made by manually reading the text and assigning the correct concept [28,29]
- Rule-based: the mapping is made using rule-based methods such as text search, regular expressions, finite state machines, or a tool that is defined as rule-based [30,31]
- Machine learning: the mapping is made using probabilistic algorithms based on a learning mechanism such as support vector machine, conditional random fields [32], or naïve Bayes [33]
- Hybrid: the mapping is made using both rule-based and machine learning methods, whether it is simultaneously combined or sequentially [34]

Results

Article Selection

After 3 rounds of recursive reference review, the final selection included 76 publications and 3 reviews. Complete list of the publications is provided in Multimedia Appendix 1 [14,16,28-101]. Those reviews [13,102,103] will be excluded from the rest of the analysis, as they were only studied to broaden the scope of this review. The flow diagram according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [104] is shown in Figure 1.

Among the 76 selected articles, 42 (55%) publications were journal articles and 34 (45%) were conference proceedings. The number of publications published per year is shown in Figure 2. The 76 publications were issued from 37 journals and conference proceedings, with 10 journals or proceedings appearing in more than one publication in the selection (Table 1).

Overall, 238 unique authors were credited in the selection. More prolific authors (more than one authorship) are displayed in Figure 3.

Figure 1. Flow diagram of the selection process. SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

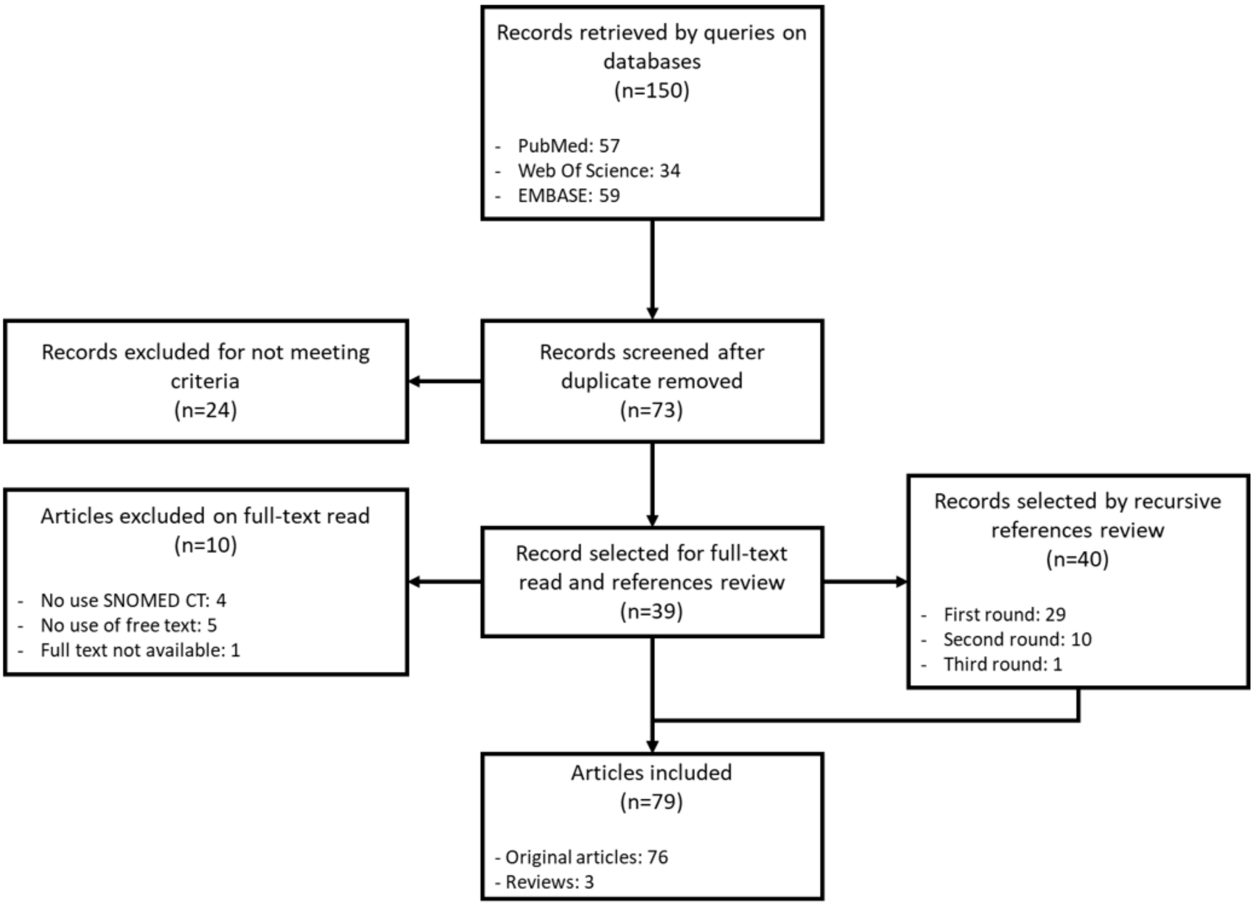


Figure 2. Number of publications per year of publication.

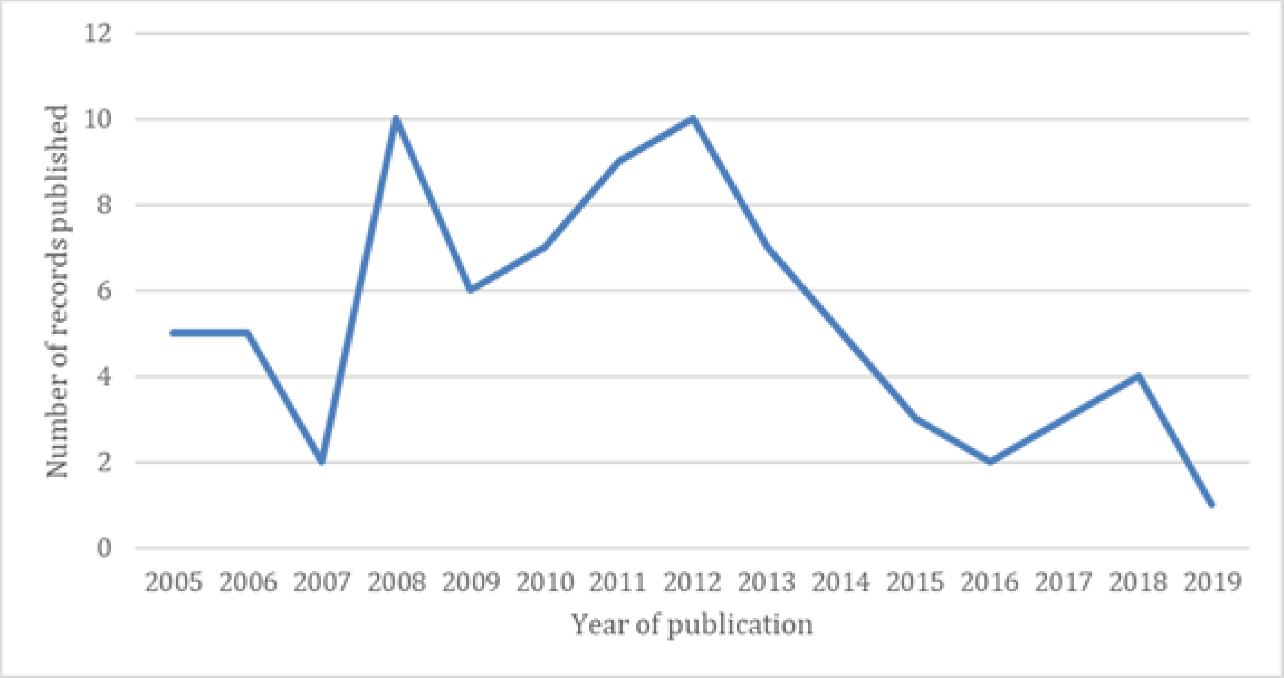


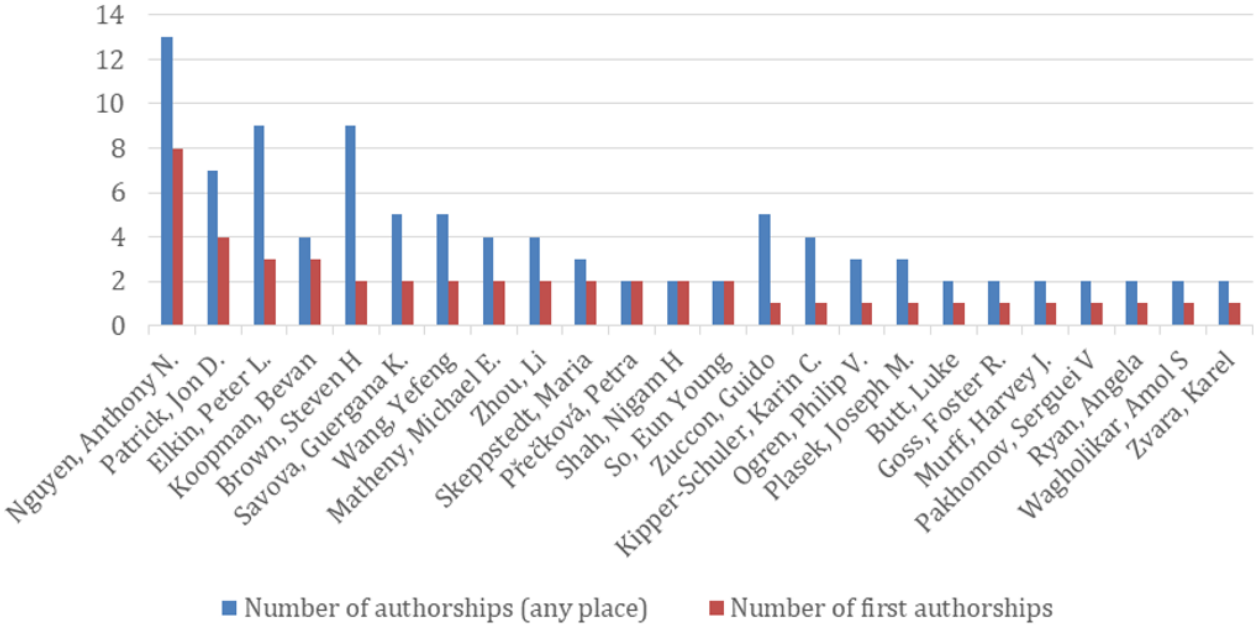
Table 1. Journals and conferences having more than 1 article in the selection.

Name of journal or conference	Publications, n
AMIA[a] *Annual Symposium proceedings*	15
Journal of Biomedical Informatics	8
BMC[b] *Medical Informatics and Decision Making*	7
Journal of the American Medical Informatics Association	7
Studies in Health Technology and Informatics	3
Journal of Digital Imaging	2
AMIA Joint Summits on Translational Science proceedings	2
Mayo Clinic Proceedings	2
Electronic Journal of Health Informatics	2
International Journal of Medical Informatics	2

[a] AMIA: American Medical Informatics Association.

[b] BMC: BioMed Central.

Figure 3. Number of authorships for the most prolific authors in selection.



Type of Data

The types of documents used in each publication are summarized in Table 2. The most frequent types are complementary exam reports (18/76, 24%), followed by narrative notes (16/76, 21%) and publications using more than one type of document (14/76, 18%).

Table 2. Number of publications per type of document used for the mapping.

Document Type	Publications (N=76), n (%)
Complementary exam report	18 (24)
Narrative note	16 (21)
Various	14 (18)
History and physical examination	8 (11)
Clinical summary	6 (8)
Death certificate	5 (7)
Problem list	3 (4)
Not available	3 (4)
Incident report	1 (1)
Autopsy report	1 (1)
Allergy report	1 (1)

Language

The target languages in the publications are listed in Table 3. Most papers focused on English (69/76, 91%). The 3 other languages were Swedish, Czech, and Chinese (Table 3).

Table 3. Target language in publications.

Language	Publications (N=76), n (%)
English	69 (91)
Swedish	3 (4)
Czech	3 (4)
Chinese	1 (1)

Reason for the Terminology Binding to SNOMED CT

As the focus of this work is to depict how the research community uses SNOMED CT to process clinical free text, selected articles had to include a part in which free-text data were mapped to SNOMED CT concepts. However, the mapping part was only a step toward another goal in many cases (eg, classification task [35,36], similarity measures [29,37], etc; Table 4).

Table 4. Role of the Systematized Nomenclature of Medicine Clinical Terms mapping in the publications.

Role of the SNOMED CT[a] mapping	Publications (N=76), n (%)
Final goal	16 (21)
Part of final goal	25 (33)
Step toward other goal	35 (46)

[a] SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

The reasons for the SNOMED CT mapping in publications are displayed in Table 5. The most frequent reason is information extraction (44/76, 39%), followed by feature in a classification task (26/76, 23%) and data normalization (23/76, 20%). The remaining categories appear in 5 publications or less.

Table 5. Reason for the mapping in publications.

Reason for the SNOMED CT[a] mapping	Publications, n (%)
Information extraction	44 (39)
Feature in a classification task	26 (23)
Data normalization	23 (20)
Coverage evaluation	5 (4)
Similarity evaluation	4 (4)
Quality evaluation	3 (3)
Value set creation	3 (3)
Synonym resource	2 (2)
Terminology mapping	2 (2)
Gold standard creation	1 (1)
Total number of points given	113 (100)

[a] SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

Mapping Method

The type of method used for mapping according to the previously defined classification is presented in Table 6, and the methods used per year is displayed in Figure 4. The evolution of the methods shows that articles presenting machine learning approaches were published only in 2008, 2009, and 2014. Hybrid approaches are present during the period 2005 to 2010 and in 2019. With 70% (53/76) of publications, rule-based approaches were the most common method used to perform this task, although the number of publications per year is reducing overall.

Table 6. Method used for mapping free-text data to Systematized Nomenclature of Medicine Clinical Terms.

Method for SNOMED CT[a] mapping	Publications (N=76), n (%)
Rule-based	53 (70)
Manual	11 (14)
Hybrid	8 (11)
Machine learning	4 (5)

[a] SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

Figure 4. Number of articles applying a specific method for Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) mapping when available.

Software Used for Terminology Binding

Table 7 shows the software used to specifically map free-text data to SNOMED CT concepts, the number of publications in which they appear, and whether they are publicly available. Only software used to produce a mapping into SNOMED CT are considered. Software used only for a step of the NLP pipeline such as negation detection or tokenization and not resulting in a concept-mapping output are not listed.

Table 7. Tools used for mapping free text to Systematized Nomenclature of Medicine Clinical Terms concepts.

Name of tool	Publications, n	Availability for public use
Medtex	12	No
MCVS[a]	8	No
MTERMS[b]	4	No
MetaMap	3	Yes
MetaMap transfer	3	Yes
Open biomedical annotator	2	Yes
MedLEE[c]	2	No
cTAKES[d]	2	Yes
Lingoengine	1	Yes
Snapper	1	No
iSCOUT	1	No
RapTAT[e]	1	No

[a] MCVS: Mayo Clinic Vocabulary Server.

[b] MTERMS: Medical Text Extraction Reasoning and Mapping System.

[c] MedLEE: Medical Language Extraction and Encoding System.

[d] cTAKES: clinical Text Analysis and Knowledge Extraction System.

[e] RapTAT: Rapid Text Annotation Tool.

Although all the software aim to detect concepts in free text, the wide disparities in methods and evaluation metrics, the subsets of concepts used, and the output terminologies prevent strict comparison. Therefore, the following review focuses only on the systems themselves and their published evaluation.

Medtex [38]**,** developed by the Australian eHealth research center, is built based on other existing tools (GATE [105], metamap transfer MMTx [106], and NegEx [107]) and can annotate free text with SNOMED CT concepts and negation marks. Although it is used in 12 publications, to the best of our knowledge, no strict evaluation of the mapping has been published.

The *Mayo Clinic Vocabulary Server* (MCVS) [16], also called Multi-threaded Clinical Vocabulary Server [39], is able to map free text to codes in various classifications, among which, SNOMED CT codes. It is the subject of an evaluation of over 4996 problem statements, which resulted in a sensitivity of 99.7% and a specificity of 97.9%. It is linked to *LingoEngine* [40], which is described as a commercially available product linked to MCVS.

The *Medical Text Extraction Reasoning and Mapping System (MTERMS)* [41] is a system that uses shallow and deep parsers to extract and structure information from free text by using local and standard terminologies. The system also proposes mappings between the terminologies. It has been used to extract medication information, allergens, allergic reactions, [42,43] and family relatives [44]. Each of these uses required specific customization, such as adding ad hoc dictionaries. Evaluations proposed in publications about MTERMS cover the encoding of information in multiple terminologies and are restricted to a specific subject. The evaluation of allergy data shows a precision of 84.4%, a recall of 91.0%, and an F-measure of 87.6%. Moreover, the evaluation of family relatives showed a precision of 100%, a recall of 97.4%, and an F-measure of 98.7% over 291 occurrences.

MetaMap [106]**,** and its Java implementation *MMTx*, was developed by the NLM. Its goal is to map the biomedical text to the UMLS Metathesaurus [108]. Since 2004, the UMLS Metathesaurus contains SNOMED CT. Although MetaMap only maps free text to the UMLS concept unique identifier(CUI), the link between a CUI and a SNOMED CT concept is present in the Metathesaurus and it is possible to specify vocabulary sources used for mapping. Therefore, in this work, MetaMap is considered as a tool that can map free text to SNOMED CT concepts. A realistic evaluation of the performance of this software has never been performed [109]. However, specific task evaluations and comparisons with other software have been published [110-112]. They showed a performance of 88% in recall, 89% in precision, and 88% in F-score on clinical notes; a precision of 85% and a recall of 78% on concepts extracted from medical curriculum documents [110]; and finally, a precision between 33% and 76% on multiple web-based biomedical resources for the mapping of biological processes, depending on the data sources [111]. However, no specific evaluation of the SNOMED CT mapping has been published.

The *Open Biomedical Annotator* (*OBA*) [113] is an ontology-based web service that can annotate free text with a variety of ontologies. It uses and improves the annotations of a concept recognizer called Mgrep [114] and is developed by the National Center for Integrative Biomedical Informatics at the University of Michigan. Publications using OBA in the selection did not propose an evaluation of the SNOMED CT mapping. However, a comparison of Mgrep with Metamap showed a precision between 58% and 93% for biological processes depending on the data source [111]. However, these evaluations are not focused on SNOMED CT.

The *Medical Language Extraction and Encoding System* [115] developed in Columbia University aims to transform clinical data into controlled vocabularies. It has been specifically adapted for UMLS and evaluated on 300 random sentences with a precision of 89% and a recall of 83% [116]. However, this evaluation does not mention SNOMED CT or the UMLS version used.

The *clinical Text Analysis and Knowledge Extraction System* (*cTAKES*) [45], developed in the Mayo Clinic, is an open-source NLP software aimed at information extraction. It includes a dictionary lookup component able to map the free-text data to UMLS concepts. The named entity recognition component has been evaluated on a corpus of 160 notes manually annotated with UMLS concepts including SNOMED CT, and shows an F-score of 71.5% for exact and 82.4% for overlapping spans [46].

Snapper [117] by the Australian eHealth research center is a software with the ability to input free-text data and perform the mapping from a terminology to SNOMED CT. To the best of our knowledge, no strict evaluation of the software has been performed. *Snapper* has been used in the selection to classify narratives into symptom groups [47].

ISCOUT appears in only one publication in the selection. This software, developed at the Brigham and Women's Hospital in Boston, is used internally for document retrieval according to a list of terms from a terminology [48]. In the publication, it is used with a list of concepts from various terminologies, including SNOMED CT, to retrieve documents. However, no evaluation of concept detection is proposed.

The *Rapid Text Annotation Tool* (*RapTAT*) [33] is a token order–specific naïve Bayes–based machine learning system designed to predict an association between phrases and concepts. It has been evaluated on the manually annotated 2010 i2b2 shared task data [118] and compared with the MCVS output, defined as the gold standard on 2860 discharge summaries. On the manual data set, *RapTAT* reached a precision of 95%, a recall of 96%, and an F-measure of 95%. To reproduce the MCVS output, *RapTAT* achieved a precision of 92%, a recall of 85%, and an F-measure of 89%.

Among all software, 5 are available, either as a web-based interface or as an installer for public usage. For example, Metamap, MMTx, and cTAKES are open source, OBA is available as a web-based interface, and LingoEngine is commercially available.

Subset Usage

As SNOMED CT includes more than 340,000 concepts, the research studies described in publications often restrict their usage to a subset of the terminology (Table 8). The complete SNOMED CT terminology was used in 64% (49/76) of the publications. A subset of the terminology was used in 30%

(23/76). The size of these subsets could vary from less than 10 concepts [47] to several thousand [37].

Table 8. Subset of Systematized Nomenclature of Medicine Clinical Terms used in publications.

Subset of SNOMED CT[a] used	Publications (N=76), n (%)
Full terminology	49 (64)
Subset	23 (30)
Not available	4 (5)

[a] SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

Advanced Functionalities Used

SNOMED CT includes a large set of functionalities atop the classical ontology usage, among which the most interesting are the combinatorial possibilities that offer postcoordination. Table 9 shows whether a publication performed postcoordination to a certain extent. Among the 13 publications using this feature, 4 of them (5%) [30,35,36,49], all by the same first author, specifically mentioned the compositional grammar published by SNOMED CT [10]; however, the others do not elaborate nor propose simple postcoordination such as combining concepts with a "+" sign.

Table 9. Use of postcoordination.

Usage of postcoordination	Publications (N=76), n (%)
No	61 (80)
Yes	13 (17)
Not available	2 (3)

Discussion

Principal Findings

SNOMED CT is mostly used to represent information found in the complementary exam reports (18/796, 24%). This is potentially influenced by an important number of studies focusing on radiology [119] and pathology, as complementary exam reports are often produced by those divisions. Moreover, pathology being historically the field of SNOMED CT, it could have influenced its application in this domain. In addition, these types of reports are usually focused on specific clinical questions and arguably convey more specific informational content.

The second type of free text represented in our results is narrative notes (16/76, 21%). Potentially, this can be explained by the large conceptual span of SNOMED CT, which allows good informational coverage on textual data.

Finally, a large set of articles do not filter data for specific types. This is explained by publications focusing more on providing a solution to map SNOMED CT concepts to text in general, without targeting a specific type of document. This is supported by the fact that those publications have the mapping to SNOMED CT concepts as their final goal in 9 out of 14 (64%) publications, which is significantly higher than the rest of the selection (16/76, 21%).

In the selection, only 7 out of 76 (9%) publications focused on a language other than English. Multiple reasons can explain this predominance of the English language in research studies. First, NLP is known to be dependent on language. Work performed in a language cannot easily be transferred to other languages. Therefore, the overhead to begin NLP research in another language is substantial and brings few rewards in the first stages, as the breakthrough has already been published in another language.

Second, SNOMED CT—like most international classifications and ontologies—was first published only in English. Rule-based methods, which are the most frequently used methods for SNOMED CT mapping, rely on the assumption that the description of a concept can be directly mapped to free text, which is not possible when the language of the text is not the language of the classification. However, translations of SNOMED CT exist for Spanish, Swedish, and recently French [120]. Therefore, there is hope for new developments as the barriers to the language start to be overcome.

Finally, several publications use public data sets such as the i2b2-shared task data sets [33,34,41,50,51] or the MIMIC II [52] data set as the sources of narrative documents. These public data sets are valuable for promoting research in NLP on clinical free text and are the subject of many publications. The availability of such resources in languages other than English is scarce.

Unsurprisingly, the most frequent reason for mapping to SNOMED CT is information extraction (44/76, 39%), as the ability of SNOMED CT to represent medical knowledge is the core feature of this terminology. Nonetheless, 26 articles (34.21%) used the resulting mapping as a feature in a classification task, usually using a learning algorithm such as support vector machines or conditional random fields [53,54]. SNOMED CT is used in these cases as a proxy for the semantic content of the data, between free text and structured data, to simplify the task of classification and improve results.

Similarity evaluation is the goal in 4 publications (5.26%). Whether it is to compare cases [55], documents [29,37], or concepts [56], the similarity is computed using the SNOMED

CT concepts. Both the polyhierarchy and the defining relationships can be used to compute the semantic distance between concepts. However, only 3 of the publications used them. This is an example of the added value SNOMED CT can bring to the secondary use of medical data.

Only 21% (16/76) of the publications mapped free text to SNOMED CT as a final objective. This is explained by the large number of publications reusing a mapping tool developed in a previous publication for new goals. To illustrate this phenomenon, Nguyen et al [38] reuse the software Medtex presented their study in multiple publications [14,30,35,36,49]. This is also true for large publicly available tools such as MCVS [16,17,57] or MTERMS [41,42].

The 3 most represented software in the selection—Medtex, MTERMS, and MCVS—are not available for public use. They mainly appear in publications by teams that have developed them. However, 2 software packages are available under an open-source license and can be freely used to map free text to SNOMED CT concepts, Metamap (and MMTx), and cTAKES. These tools are available to perform automatic annotation with SNOMED CT; however, none of them are specifically aimed at this ontology nor do they include features such as postcoordination or multiple language support. There is currently no clear solution for mapping free text to SNOMED CT concepts out of the box with a specific focus on this ontology and its features. This could explain the overall small number of publications in the selection.

Rule-based methods are largely used to perform mapping (53/76, 70%). This tends to show that they are more suited for this task. This phenomenon could be due to the large number of concepts in SNOMED CT. The amount of annotated data needed to automatically map free text with more than 340,000 classes is enormous and would require an important investment.

The evaluations of the automatic mapping found in publications show that this is not a trivial task. Most solutions for mapping lack a clear and definitive evaluation, and when available, they usually focus on a small set of documents; they use a subset of the terminology or do not rely on a gold standard. This gap in research could be explained by several reasons.

The number of concepts in SNOMED CT is large, and all granularities coexist. To express a simple concept such as *Tuberculous pneumonia*, a single concept can be used: 80003002 (Tuberculous pneumonia [disorder]) or any combination of less granular concepts (233604007 | Pneumonia [disorder], 233618000 |Mycobacterial pneumonia [disorder], 56717001 | Tuberculosis [disorder], 113858008 |Mycobacterium tuberculosis complex [organism], etc). However, all these representations can be equally correct from a semantic point of view. Therefore, it is difficult to compute the recall as a gold standard, which usually represents only one of these representations. Moreover, SNOMED CT contains 18 subhierarchies focusing on different thematics (clinical findings, body structure, etc), which make the decision of which concept to use even more difficult. For example, the hierarchy of the observable entities defines what can be observed in a patient, but the clinical finding hierarchy contains the results of those observations. The choice between a finding and an observable entity is not always clear and can heavily depend on the context. Finally, the usage of postcoordinated terms increases the set of expressions that can be used to represent the same concept. Overall, the task of evaluating the automatic mapping of natural language to a SNOMED CT concept lacks a pragmatic and applicable method; therefore, it is often limited to small-scale evaluations or manual validations.

The version of SNOMED used in publications (SNOMED, SNOMED CT, or SNOMED RT) is not always specified, especially when the usage of this terminology is not the main goal of the research. Moreover, the usage of SNOMED CT is implicit when UMLS is used. This remark, as well as the small number of publications mentioning postcoordination, emphasizes the fact that SNOMED CT is often seen as a simple terminology, without the need to use its advanced features. This phenomenon is also shown by the fact that only a subset of the terminology is used in 64% (49/76) of the publications. Using a subset simplifies the mapping task by reducing complexity but also prevents from benefiting from the power of the polyhierarchy and the relationships among concepts.

As clinical free text is written in natural language and since SNOMED CT is designed as a formal language, it is surprising that very few papers use this functionality when mapping to free text. Although this can be explained by the fact that even if SNOMED International provides compositional grammar, there is, to the best of our knowledge, no explicit roadmap to use it for such a task. Postcoordination requires deep knowledge of the terminology and access to a terminology server that handles the resulting data. As SNOMED International is not a software provider, this has to be achieved either using the open-source server Snowstorm [121], for which SNOMED International does not provide technical support, or by relying on a private company software.

This work shows that although SNOMED CT is widely used in health care, its use to represent free-text data still remains a challenge. Polyhierarchy and compositional grammar are at the core of SNOMED CT and they can bring significant value to data; however, when it comes to mapping concepts to free text, there seems to be a margin for approaches that take advantage of those features. The same can be observed on the usage of SNOMED CT to process free text in languages other than English.

Although machine learning is clearly on the rise in multiple fields of medical informatics and scientific research in general, it is rarely used to map free text to SNOMED CT, most probably because of the size of the corpus needed to train on such a large set of classes. In contrast, rule-based symbolic approaches seem more suited and are used to map large terminologies to free-text data. A combination of the strengths of both hybrid approaches could be a way to improve performance.

Finally, an openly available tool that would process free texts and map them to SNOMED CT concepts is yet to be created.

Limitations

Although the review has been conducted following a systematic approach, this work has some limitations.

The last publication research was conducted in August 2019. It is possible that new publications have been published since then. As we have observed, the number of publications selected per year is reducing; therefore, we consider the impact of this gap to be arguably small. Although the recursive reference review has been performed with the aim of broadening the scope of the included papers, it is possible that some studies that have not yet been cited by other papers have not been considered. For example, the high-throughput phenotyping NLP system described by Schlegel et al. [122] did not appear in the search nor during the recursive reference review. This system uses a series of linguistic and semantic indexes to process clinical data and characterizes it using ontologies such as SNOMED CT and the International Classification of Diseases 10.

In the selection, a large number of publications are published by the same groups of authors and propose similar works. This could result in an overestimation of the impact of those publications on a complete selection.

Finally, it is possible that because of the choice to focus on biomedical databases to gather publications, some articles published on more engineering-oriented databases have not been included.

Conclusions

In conclusion, clinical free-text processing and SNOMED CT have been an important subject for research, but the number of publications has been diminishing in recent years. Most of the publications that we found mapped free text to SNOMED CT to obtain a semantic representation of the data and used it as a first step toward other goals such as document classification or information retrieval.

Almost none of the publications used advanced features of SNOMED CT, such as the polyhierarchy or postcoordination. Most publications conceive SNOMED CT only as a terminology, a dictionary, or a resource for synonyms.

Publications focusing on languages other than English are rare and, if software exists for mapping English free text to SNOMED CT, most of them are not available for public use or focus on UMLS and not strictly on SNOMED CT. There is currently no easy solution for mapping free-text data into the SNOMED CT concepts, especially if the source language is different from English or if postcoordination is needed.

However, the need for formal semantic representation of health care data and the secondary use of free-text data is high, and automatic encoding into a compositional ontology could be a way to achieve interoperability.

Acknowledgments

This research was funded by the Language and Communication Network of the University of Geneva.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of 76 articles included in the review process.
[PDF File (Adobe PDF File), 507 KB - jmir_v23i1e24594_app1.pdf]

References

1. Kierkegaard P. E-prescription across Europe. Health Technol 2012 Dec 20;3(3):205-219. [doi: 10.1007/s12553-012-0037-0]
2. Lau L, Shakib S. Towards Data Interoperability: Practical Issues in Terminology Implementation and Mapping. In: HIC 2005 and HINZ 2005: Proceedings. HIC 2005 and HINZ 2005: Proceedings Health Informatics Society of Australia; 2005 Presented at: HIC 2005: Thirteenth National Health Informatics Conference ; HINZ 2005: Fourth Health Informatics Conference; 2005; Brunswick East, Vic.
3. Grimson J, Murphy J. The Jupiter approach to interoperability with healthcare legacy systems. Medinfo 1995;8 Pt 1:367-371. [Medline: 8591200]
4. Randorff Højen A, Rosenbeck Gøeg K. Snomed CT implementation. Mapping guidelines facilitating reuse of data. Methods Inf Med 2012;51(6):529-538. [doi: 10.3414/ME11-02-0023] [Medline: 23038162]
5. SNOMED CT Starter Guide. SNOMED Confluence. URL: https://confluence.ihtsdotools.org/display/DOCSTART/SNOMED+CT+Starter+Guide?preview=/28742871/47677485/doc_StarterGuide_Current-en-US_INT_20170728.pdf [accessed 2019-06-14]
6. SNOMED. URL: http://www.snomed.org/ [accessed 2019-06-14] [WebCite Cache ID http://www.snomed.org/]
7. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. J Am Med Inform Assoc 2014 Feb;21(e1):e11-e19 [FREE Full text] [doi: 10.1136/amiajnl-2013-001636] [Medline: 23828173]
8. Members. SNOMED. URL: https://www.snomed.org/our-customers/members [accessed 2020-09-01]
9. Agrawal A, He Z, Perl Y, Wei D, Halper M, Elhanan G, et al. The readiness of SNOMED problem list concepts for meaningful use of electronic health records. Artif Intell Med 2013 Jun;58(2):73-80. [doi: 10.1016/j.artmed.2013.03.008] [Medline: 23602702]

10. SNOMED CT Compositional Grammar. SNOMED Confluence. URL: https://confluence.ihtsdotools.org/display/SLPG/SNOMED+CT+Compositional+Grammar [accessed 2020-09-01]
11. SNOMED CT concept model. SNOMED. URL: https://confluence.ihtsdotools.org/display/DOCGLOSS/SNOMED+CT+concept+model [accessed 2020-09-01]
12. Cohen K, Demner-Fushman D. Biomedical Natural Language Processing. Amsterdam: John Benjamins Publishing Company; 2014:978-990.
13. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform 2008:128-144. [Medline: 18660887]
14. Zuccon G, Wagholikar AS, Nguyen AN, Butt L, Chu K, Martin S, et al. Automatic Classification of Free-Text Radiology Reports to Identify Limb Fractures using Machine Learning and the SNOMED CT Ontology. AMIA Jt Summits Transl Sci Proc 2013;2013:300-304 [FREE Full text] [Medline: 24303284]
15. Chapman WW, Nadkarni PM, Hirschman L, D'Avolio LW, Savova GK, Uzuner O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. J Am Med Inform Assoc 2011 Oct;18(5):540-543 [FREE Full text] [doi: 10.1136/amiajnl-2011-000465] [Medline: 21846785]
16. Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc 2006 Jun;81(6):741-748. [doi: 10.4065/81.6.741] [Medline: 16770974]
17. Garvin JH, Elkin PL, Shen S, Brown S, Trusko B, Wang E, et al. Automated quality measurement in Department of the Veterans Affairs discharge instructions for patients with congestive heart failure. J Healthc Qual 2013;35(4):16-24. [doi: 10.1111/j.1945-1474.2011.195.x] [Medline: 23819743]
18. Bakhshi-Raiez F, de Keizer NF, Cornet R, Dorrepaal M, Dongelmans D, Jaspers MW. A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. Int J Med Inform 2012 May;81(5):351-362. [doi: 10.1016/j.ijmedinf.2011.09.010] [Medline: 22030036]
19. Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. AMIA Annu Symp Proc 2003:699-703 [FREE Full text] [Medline: 14728263]
20. Cornet R, de KN. Forty years of SNOMED: a literature review. BMC Med Inform Decis Mak 2008 Oct 27;8 Suppl 1:S2 [FREE Full text] [doi: 10.1186/1472-6947-8-S1-S2] [Medline: 19007439]
21. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 2016 Feb 05;23(5):1007-1015 [FREE Full text] [doi: 10.1093/jamia/ocv180]
22. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2020-09-01]
23. Embase. URL: https://www.embase.com/#search [accessed 2020-09-01]
24. Web of Science. URL: https://apps.webofknowledge.com [accessed 2019-06-14]
25. Lindberg DA, Humphreys BL, McCray AT. The unified medical language system. Yearb Med Inform 2018 Mar 05;02(01):41-51. [doi: 10.1055/s-0038-1637976]
26. 2004AA UMLS Documentation. US National Library of Medicine. URL: https://www.nlm.nih.gov/archive/20080407/research/umls/archive/2004AA/UMLSDOC.html [accessed 2020-09-01]
27. US National Library of Medicine. National Institutes of Health (NIH). URL: https://www.nlm.nih.gov/research/umls/Snomed/snomed_represented.html [accessed 2020-09-01]
28. So E, Park H. Mapping medical records of gastrectomy patients to SNOMED CT. Stud Health Technol Inform 2011;169:764-768. [Medline: 21893850]
29. Přečková P, Zvárová J, Zvára K. Measuring diversity in medical reports based on categorized attributes and international classification systems. BMC Med Inform Decis Mak 2012 Apr 12;12:31 [FREE Full text] [doi: 10.1186/1472-6947-12-31] [Medline: 22498343]
30. Nguyen A, Lawley M, Hansen D, Colquist S. Structured pathology reporting for cancer from free text: Lung cancer case study. In: HIC 2010: Proceedings. 2012 Presented at: 18th Annual Health Informatics Conference: Informing the Business of Healthcare; August 24-26, 2010; Melbourne URL: https://search.informit.com.au/documentSummary;dn=429807729626920;res=IELHEA;type=pdf
31. Patrick J, Wang Y, Budd P. An automated system for conversion of clinical notes into SNOMED clinical terminology. In: Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68. 2007 Presented at: Australasian symposium on ACSW frontiers; January 30 - February 2, 2007; Ballarat, Victoria, Australia p. 226. [doi: 10.1007/978-1-4471-2801-4_13]
32. Wang Y, Patrick JD. Cascading Classifiers for Named Entity Recognition in Clinical Notes. In: WBIE '09: Proceedings of the Workshop on Biomedical Information Extraction. 2009 Presented at: WBIE: Workshop on Biomedical Information Extraction; September 2009; Bulgaria p. 42 URL: https://www.aclweb.org/anthology/W09-4507
33. Gobbel GT, Reeves R, Jayaramaraja S, Giuse D, Speroff T, Brown SH, et al. Development and evaluation of RapTAT: a machine learning system for concept mapping of phrases from medical narratives. J Biomed Inform 2014 Apr;48:54-65 [FREE Full text] [doi: 10.1016/j.jbi.2013.11.008] [Medline: 24316051]

34. Patrick J, Li M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. J Am Med Inform Assoc 2010 Oct;17(5):524-527 [FREE Full text] [doi: 10.1136/jamia.2010.003939] [Medline: 20819856]
35. Nguyen A, Moore J, Zuccon G, Lawley M, Colquist S. Classification of pathology reports for cancer registry notifications. Stud Health Technol Inform 2012;178:150-156. [Medline: 22797034]
36. Nguyen AN, Lawley MJ, Hansen DP, Bowman RV, Clarke BE, Duhig EE, et al. Symbolic rule-based classification of lung cancer stages from free-text pathology reports. J Am Med Inform Assoc 2010;17(4):440-445 [FREE Full text] [doi: 10.1136/jamia.2010.003707] [Medline: 20595312]
37. Mabotuwana T, Lee MC, Cohen-Solal EV. An ontology-based similarity measure for biomedical data-application to radiology reports. J Biomed Inform 2013 Oct;46(5):857-868 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.013] [Medline: 23850839]
38. Nguyen A, Lawley M, Hansen D, Colquist S. A Simple Pipeline Application for Identifying and Negating SNOMED Clinical Terminology in Free Text. In: HIC 2009: Proceedings; Frontiers of Health Informatics - Redefining Healthcare. 2009 Presented at: HIC 2009: Frontiers of Health Informatics; August 19-21, 2009; National Convention Centre Canberra, Australia p. 196. [doi: 10.1007/978-1-84882-803-2_11]
39. Elkin A, Froehling D, Wahner-Roedler D, Trusko B, Welsh G, Ma H, et al. NLP-based identification of pneumonia cases from free-text radiological reports. AMIA Annu Symp Proc 2008 Nov 06:172-176 [FREE Full text] [Medline: 18998791]
40. Brown SH, Elkin PL, Rosenbloom ST, Fielstein E, Speroff T. eQuality for all: extending automated quality measurement of free text clinical narratives. AMIA Annu Symp Proc 2008:71-75 [FREE Full text] [Medline: 18999230]
41. Zhou L, Plasek JM, Mahoney LM, Karipineni N, Chang F, Yan X, et al. Using Medical Text Extraction, Reasoning and Mapping System (MTERMS) to process medication information in outpatient clinical notes. AMIA Annu Symp Proc 2011;2011:1639-1648 [FREE Full text] [Medline: 22195230]
42. Goss FR, Plasek JM, Lau JJ, Seger DL, Chang FY, Zhou L. An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. AMIA Annu Symp Proc 2014;2014:580-588 [FREE Full text] [Medline: 25954363]
43. Plasek JM, Goss FR, Lai KH, Lau JJ, Seger DL, Blumenthal KG, et al. Food entries in a large allergy data repository. J Am Med Inform Assoc 2016 Apr;23(e1):e79-e87 [FREE Full text] [doi: 10.1093/jamia/ocv128] [Medline: 26384406]
44. Zhou L, Lu Y, Vitale CJ, Mar PL, Chang F, Dhopeshwarkar N, et al. Representation of information about family relatives as structured data in electronic health records. Appl Clin Inform 2014;5(2):349-367 [FREE Full text] [doi: 10.4338/ACI-2013-10-RA-0080] [Medline: 25024754]
45. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
46. Ogren P, Savova G, Chute C. Constructing evaluation corpora for automated clinical named entity recognition. In: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems. Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems IOS Press; 2007 Presented at: Medinfo 2007; 2007; Brisbane, Australia p. 2325 URL: https://www.aclweb.org/anthology/L08-1366/
47. Wagholikar A, Lawley MJ, Hansen DP, Chu K. Identifying symptom groups from Emergency Department presenting complaint free text using SNOMED CT. AMIA Annu Symp Proc 2011;2011:1446-1453 [FREE Full text] [Medline: 22195208]
48. Warden GI, Lacson R, Khorasani R. Leveraging terminologies for retrieval of radiology reports with critical imaging findings. AMIA Annu Symp Proc 2011;2011:1481-1488 [FREE Full text] [Medline: 22195212]
49. Nguyen A, Moore J, Lawley M, Hansen D, Colquist S. Automatic extraction of cancer characteristics from free-text pathology reports for cancer notifications. Stud Health Technol Inform 2011;168:117-124. [Medline: 21893919]
50. Patrick JD, Nguyen DH, Wang Y, Li M. A knowledge discovery and reuse pipeline for information extraction in clinical notes. J Am Med Inform Assoc 2011;18(5):574-579 [FREE Full text] [doi: 10.1136/amiajnl-2011-000302] [Medline: 21737844]
51. Jindal P, Roth D. Extraction of events and temporal expressions from clinical narratives. J Biomed Inform 2013 Dec;46 Suppl:S13-S19 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.010] [Medline: 24022023]
52. Henriksson A, Conway M, Duneld M, Chapman WW. Identifying synonymy between SNOMED clinical terms of varying length using distributional analysis of electronic health records. AMIA Annu Symp Proc 2013;2013:600-609 [FREE Full text] [Medline: 24551362]
53. Li D, Savova G, Schuler K, Kipper-Schuler K, Savova G, Schuler K. Conditional Random Fields and Support Vector Machines for Disorder Named Entity Recognition in Clinical Texts. In: Proceedings of the workshop on current trends in biomedical natural language processing. 2008 Presented at: BioNLP 2008: Current Trends in Biomedical Natural Language Processing; 2008; Columbus, Ohio p. 94. [doi: 10.3115/1572306.1572326]
54. Aseervatham S, Bennani Y. Semi-structured document categorization with a semantic kernel. Pattern Recognition 2009 Sep;42(9):2067-2076. [doi: 10.1016/j.patcog.2008.10.024]

55. Melton GB, Parsons S, Morrison FP, Rothschild AS, Markatou M, Hripcsak G. Inter-patient distance metrics using SNOMED CT defining relationships. J Biomed Inform 2006 Dec;39(6):697-705 [FREE Full text] [doi: 10.1016/j.jbi.2006.01.004] [Medline: 16554186]
56. Pivovarov R, Elhadad N. A hybrid knowledge-based and data-driven approach to identifying semantically similar concepts. J Biomed Inform 2012 Jun;45(3):471-481 [FREE Full text] [doi: 10.1016/j.jbi.2012.01.002] [Medline: 22289420]
57. Matheny ME, FitzHenry F, Speroff T, Green JK, Griffith ML, Vasilevskis EE, et al. Detection of infectious symptoms from VA emergency department and primary care clinical documentation. Int J Med Inform 2012 Mar;81(3):143-156. [doi: 10.1016/j.ijmedinf.2011.11.005]
58. Tahmasebi AM, Zhu H, Mankovich G, Prinsen P, Klassen P, Pilato S, et al. Automatic normalization of anatomical phrases in radiology reports using unsupervised learning. J Digit Imaging 2019 Feb;32(1):6-18 [FREE Full text] [doi: 10.1007/s10278-018-0116-5] [Medline: 30076490]
59. Jackson R, Patel R, Velupillai S, Gkotsis G, Hoyle D, Stewart R. Knowledge discovery for deep phenotyping serious mental illness from electronic mental health records. F1000Res 2018;7:210 [FREE Full text] [doi: 10.12688/f1000research.13830.2] [Medline: 29899974]
60. Mujtaba G, Shuib L, Raj RG, Rajandram R, Shaikh K, Al-Garadi MA. Classification of forensic autopsy reports through conceptual graph-based document representation model. J Biomed Inform 2018 Jun;82:88-105 [FREE Full text] [doi: 10.1016/j.jbi.2018.04.013] [Medline: 29738820]
61. Koopman B, Zuccon G, Nguyen A, Bergheim A, Grayson N. Extracting cancer mortality statistics from death certificates: a hybrid machine learning and rule-based approach for common and rare cancers. Artif Intell Med 2018 Jul;89:1-9. [doi: 10.1016/j.artmed.2018.04.011] [Medline: 29754799]
62. Nguyen AN, Truran D, Kemp M, Koopman B, Conlan D, O'Dwyer J, et al. Computer-assisted diagnostic coding: effectiveness of an NLP-based approach using SNOMED CT to ICD-10 mappings. AMIA Annu Symp Proc 2018;2018:807-816 [FREE Full text] [Medline: 30815123]
63. Zvára K, Tomečková M, Peleška J, Svátek V, Zvárová J. Tool-supported interactive correction and semantic annotation of narrative clinical reports. Methods Inf Med 2017 May 18;56(3):217-229. [doi: 10.3414/ME16-01-0083] [Medline: 28451691]
64. Zhang R, Liu J, Huang Y, Wang M, Shi Q, Chen J, et al. Enriching the international clinical nomenclature with Chinese daily used synonyms and concept recognition in physician notes. BMC Med Inform Decis Mak 2017 May 02;17(1):54 [FREE Full text] [doi: 10.1186/s12911-017-0455-z] [Medline: 28464923]
65. Lin C, Hsu CJ, Lou YS, Yeh SJ, Lee CC, Su SL, et al. Artificial intelligence learning semantics via external resources for classifying diagnosis codes in discharge notes. J Med Internet Res 2017 Nov 06;19(11):e380 [FREE Full text] [doi: 10.2196/jmir.8344] [Medline: 29109070]
66. Nguyen AN, Moore J, O'Dwyer J, Philpot S. Automated cancer registry notifications: validation of a medical text analytics system for identifying patients with cancer from a state-wide pathology repository. AMIA Annu Symp Proc 2016;2016:964-973 [FREE Full text] [Medline: 28269893]
67. Nguyen AN, Moore J, O'Dwyer J, Philpot S. Assessing the utility of automatic cancer registry notifications data extraction from free-text pathology reports. AMIA Annu Symp Proc 2015;2015:953-962 [FREE Full text] [Medline: 26958232]
68. Koopman B, Karimi S, Nguyen A, McGuire R, Muscatello D, Kemp M, et al. Automatic classification of diseases from free-text death certificates for real-time surveillance. BMC Med Inform Decis Mak 2015 Jul 15;15:53 [FREE Full text] [doi: 10.1186/s12911-015-0174-2] [Medline: 26174442]
69. Koopman B, Zuccon G, Nguyen A, Bergheim A, Grayson N. Automatic ICD-10 classification of cancers from free-text death certificates. Int J Med Inform 2015 Nov;84(11):956-965. [doi: 10.1016/j.ijmedinf.2015.08.004] [Medline: 26323193]
70. Skeppstedt M, Kvist M, Nilsson GH, Dalianis H. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: an annotation and machine learning study. J Biomed Inform 2014 Jun;49:148-158 [FREE Full text] [doi: 10.1016/j.jbi.2014.01.012] [Medline: 24508177]
71. Ou Y, Patrick J. Automatic structured reporting from narrative cancer pathology reports. Electronic Journal of Health Informatics 2014;8(2).
72. Hong Y, Kahn CE. Content analysis of reporting templates and free-text radiology reports. J Digit Imaging 2013 Oct;26(5):843-849 [FREE Full text] [doi: 10.1007/s10278-013-9597-4] [Medline: 23553231]
73. So EY, Park HA. Exploring the possibility of information sharing between the medical and nursing domains by mapping medical records to SNOMED CT and ICNP. Healthc Inform Res 2011 Sep;17(3):156-161 [FREE Full text] [doi: 10.4258/hir.2011.17.3.156] [Medline: 22084810]
74. Butt L, Zuccon G, Nguyen A, Bergheim A, Grayson N. Classification of cancer-related death certificates using machine learning. Australas Med J 2013;6(5):292-299 [FREE Full text] [doi: 10.4066/AMJ.2013.1654] [Medline: 23745151]
75. Skeppstedt M, Kvist M, Dalianis H. Rule-based Entity Recognition and Coverage of SNOMED CT in Swedish Clinical Text. International Renewable Energy Conference. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/521_Paper.pdf [accessed 2020-12-28]

76. ul Muntaha S, Skeppstedt M, Kvist M, Dalianis H. Entity Recognition of Pharmaceutical Drugs in Swedish Clinical Text. 2012 Presented at: The Fourth Swedish Language Technology Conference; 2012; Lund University, Sweden p. 77-78 URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.302.5525&rep=rep1&type=pdf
77. Davis K, Staes C, Duncan J, Igo S, Facelli JC. Identification of pneumonia and influenza deaths using the Death Certificate Pipeline. BMC Med Inform Decis Mak 2012 May 08;12:37 [FREE Full text] [doi: 10.1186/1472-6947-12-37] [Medline: 22569097]
78. Liu H, Wagholikar K, Wu ST. Using SNOMED-CT to encode summary level data - a corpus analysis. AMIA Jt Summits Transl Sci Proc 2012;2012:30-37 [FREE Full text] [Medline: 22779045]
79. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. J Am Med Assoc 2011 Aug 24;306(8):848-855. [doi: 10.1001/jama.2011.1204] [Medline: 21862746]
80. Martiniz D, Li Y. Information extraction from pathology reports in a hospital setting. In: Proceedings of the 20th ACM international conference on Information and knowledge management. 2011 Presented at: 20th ACM international conference on Information and knowledge management; 2011; Glasgow, Scotland. [doi: 10.1145/2063576.2063846]
81. Fung KW, Xu J, Rosenbloom ST, Mohr D, Maram N, Suther T. Testing Three Problem List Terminologies in a simulated data entry environment. AMIA Annu Symp Proc 2011;2011:445-454 [FREE Full text] [Medline: 22195098]
82. Lee DH, Lau FY, Quan H. A method for encoding clinical datasets with SNOMED CT. BMC Med Inform Decis Mak 2010 Sep 17;10:53 [FREE Full text] [doi: 10.1186/1472-6947-10-53] [Medline: 20849611]
83. Přečková P. Language of Czech Medical Reports and Classification Systems in Medicine. European Journal of Biomedical Informatics. URL: https://www.ejbi.org/scholarly-articles/language-of-czech-medical-reports-and-classification-systems-in-medicine.pdf
84. Arnot-Smith J, Smith AF. Patient safety incidents involving neuromuscular blockade: analysis of the UK National Reporting and Learning System data from 2006 to 2008. Anaesthesia 2010 Nov;65(11):1106-1113 [FREE Full text] [doi: 10.1111/j.1365-2044.2010.06509.x] [Medline: 20840604]
85. Wang Y. Annotating and Recognising Named Entities in Clinical Notes. 2006 Presented at: Proceedings of the ACL-IJCNLP 2009 Student Research Workshop USA: Association for Computational Linguistics; 2006; The University of Sydney URL: https://www.aclweb.org/anthology/P09-3003.pdf [doi: 10.3115/1667884.1667888]
86. Matheny ME, Fitzhenry F, Speroff T, Hathaway J, Murff HJ, Brown SH, et al. Detection of blood culture bacterial contamination using natural language processing. AMIA Annu Symp Proc 2009 Nov 14;2009:411-415 [FREE Full text] [Medline: 20351890]
87. Shah NH, Jonquet C, Chiang AP, Butte AJ, Chen R, Musen MA. Ontology-driven indexing of public datasets for translational bioinformatics. BMC Bioinformatics 2009 Feb 05;10 Suppl 2:S1 [FREE Full text] [doi: 10.1186/1471-2105-10-S2-S1] [Medline: 19208184]
88. Wang Y. UIMA-based clinical information extraction system. Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP. URL: http://www.lrec-conf.org/lrec2008/IMG/ws/programme/W16.pdf [accessed 2020-12-28]
89. Schuler K, Kaggal V, Masanz J, Ogren P, Savova G. System Evaluation on a Named Entity Corpus from Clinical Notes. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08). 2008 Presented at: LREC'08; May 2008; Marrakech, Morocco URL: https://www.aclweb.org/anthology/L08-1365/
90. Ruch P, Gobeill J, Lovis C, Geissbühler A. Automatic medical encoding with SNOMED categories. BMC Med Inform Decis Mak 2008 Oct 27;8 Suppl 1:S6 [FREE Full text] [doi: 10.1186/1472-6947-8-S1-S6] [Medline: 19007443]
91. Ryan A, Patrick J, Herkes R. Introduction of enhancement technologies into the intensive care service, Royal Prince Alfred Hospital, Sydney. Health Inf Manag 2008;37(1):40-45. [doi: 10.1177/183335830803700105] [Medline: 18245864]
92. Patrick J, Wang Y, Budd P, Brandt S, Rogers B, Herkes R, et al. Developing SNOMED CT Subsets from Clinical Notes for Intensive Care Service. Health Care and Informatics Review Online. URL: https://cdn.ymaws.com/www.hinz.org.nz/resource/collection/0F09C2E4-7A05-49FB-8324-709F1AB2AA2F/F38_Patrick.pdf [accessed 2020-12-28]
93. Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, et al. An electronic health record based on structured narrative. J Am Med Inform Assoc 2008;15(1):54-64 [FREE Full text] [doi: 10.1197/jamia.M2131] [Medline: 17947628]
94. Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. J Biomed Inform 2006 Dec;39(6):589-599 [FREE Full text] [doi: 10.1016/j.jbi.2005.11.004] [Medline: 16359928]
95. Brown SH, Speroff T, Fielstein EM, Bauer BA, Wahner-Roedler DL, Greevy R, et al. eQuality: electronic quality assessment from narrative clinical reports. Mayo Clin Proc 2006 Nov;81(11):1472-1481. [doi: 10.4065/81.11.1472] [Medline: 17120403]
96. Shah NH, Rubin DL, Supekar KS, Musen MA. Ontology-based annotation and query of tissue microarray data. AMIA Annu Symp Proc 2006:709-713 [FREE Full text] [Medline: 17238433]
97. Pakhomov S, Buntrock J, Duffy P. High throughput modularized NLP system for clinical text. In: Proceedings of the ACL 2005 on Interactive poster and demonstration sessions. 2005 Presented at: ACLdemo '05; June 26, 2005; Ann Arbor, Michigan. [doi: 10.3115/1225753.1225760]
98. Long W. Extracting diagnoses from discharge summaries. AMIA Annu Symp Proc 2005:470-474 [FREE Full text] [Medline: 16779084]

99. Rindflesch TC, Pakhomov SV, Fiszman M, Kilicoglu H, Sanchez VR. Medical facts to support inferencing in natural language processing. AMIA Annu Symp Proc 2005:634-638 [FREE Full text] [Medline: 16779117]
100. Elkin PL, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR, et al. A controlled trial of automated classification of negation from clinical notes. BMC Med Inform Decis Mak 2005 May 05;5:13 [FREE Full text] [doi: 10.1186/1472-6947-5-13] [Medline: 15876352]
101. Burkhart L, Konicek R, Moorhead S, Androwich I. Mapping parish nurse documentation into the nursing interventions classification: a research method. Comput Inform Nurs 2005;23(4):220-229. [doi: 10.1097/00024665-200507000-00010] [Medline: 16027538]
102. Cornet R, Van Eldik A, De Keizer N. Inventory of tools for Dutch clinical language processing. Stud Health Technol Inform 2012;180:245-249. [Medline: 22874189]
103. Stanfill MH, Williams M, Fenton SH, Jenders RA, Hersh WR. A systematic literature review of automated clinical coding and classification systems. J Am Med Inform Assoc 2010;17(6):646-651 [FREE Full text] [doi: 10.1136/jamia.2009.001024] [Medline: 20962126]
104. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
105. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS Comput Biol 2013 Feb;9(2):e1002854 [FREE Full text] [doi: 10.1371/journal.pcbi.1002854] [Medline: 23408875]
106. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: 11825149]
107. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001 Oct;34(5):301-310 [FREE Full text] [doi: 10.1006/jbin.2001.1029] [Medline: 12123149]
108. MetaMap. URL: https://metamap.nlm.nih.gov/ [accessed 2020-09-01]
109. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
110. Denny JC, Smithers JD, Miller RA, Spickard A. 'Understanding' medical school curriculum content using KnowledgeMap. J Am Med Inform Assoc 2003;10(4):351-362 [FREE Full text] [doi: 10.1197/jamia.M1176] [Medline: 12668688]
111. Shah NH, Bhatia N, Jonquet C, Rubin D, Chiang AP, Musen MA. Comparison of concept recognizers for building the Open Biomedical Annotator. BMC Bioinformatics 2009 Sep 17;10(S9). [doi: 10.1186/1471-2105-10-s9-s14]
112. Reátegui R, Ratté S. Comparison of MetaMap and cTAKES for entity extraction in clinical notes. BMC Med Inform Decis Mak 2018 Sep 14;18(S3):-. [doi: 10.1186/s12911-018-0654-2]
113. Jonquet C, Shah NH, Musen MA. The open biomedical annotator. Summit on Translat Bioinforma 2009;2009:56-60 [FREE Full text] [Medline: 21347171]
114. Dai M, Shah NH, Xuan W, Musen MA, Watson SJ, Athey BD, et al. An efficient solution for mapping free text to ontology terms. In: AMIA summit on translational bioinformatics. 2008 Presented at: AMIA summit on translational bioinformatics; 2008; San Francisco URL: https://knowledge.amia.org/amia-55142-tbi2008a-1.650887/t-002-1.985042/f-001-1.985043/a-041-1.985157/an-041-1.985158?qr=1
115. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994;1(2):161-174 [FREE Full text] [Medline: 7719797]
116. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc 2004;11(5):392-402 [FREE Full text] [doi: 10.1197/jamia.M1552] [Medline: 15187068]
117. Vickers DM, Lawley MJ. Mapping Existing Medical Terminologies to SNOMED CT: An Investigation of the Novice User's Experience. In: HIC 2009: Proceedings; Frontiers of Health Informatics - Redefining Healthcare, National. 2009 Presented at: HIC 2009: Frontiers of Health Informatics - Redefining Healthcare; August 19-21, 2009; National Convention Centre Canberra, Australia p. 46 URL: https://cdn.ymaws.com/hisa.site-ym.com/resource/resmgr/hic2009/DVickers.pdf
118. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011 Sep 01;18(5):552-556. [doi: 10.1136/amiajnl-2011-000203]
119. Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. Radiology 2016 May;279(2):329-343. [doi: 10.1148/radiol.16142770] [Medline: 27089187]
120. An exemplar of collaboration: The first release of the SNOMED CT common French translation Internet. SNOMED. URL: http://www.snomed.org/news-and-events/articles/first-release-common-french-translation [accessed 2020-09-01]
121. IHTSDO/snowstorm. GitHub. 2020. URL: https://github.com/IHTSDO/snowstorm [accessed 2020-09-01]
122. Schlegel DR, Crowner C, Lehoullier F, Elkin PL. HTP-NLP: a new NLP system for high throughput phenotyping. Stud Health Technol Inform 2017;235:276-280. [Medline: 28423797]

Abbreviations

cTAKES: clinical Text Analysis and Knowledge Extraction System
CUI: concept unique identifier
MCVS: Mayo Clinic Vocabulary Server
MMTx: Metamap transfer
MTERMS: medical Text Extraction Reasoning and Mapping System
NLM: National Library of Medicine
NLP: natural language processing
OBA: Open Biomedical Annotator
RapTAT: Rapid Text Annotation Tool
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms
SNOMED RT: SNOMED reference terminology
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 28.09.20; peer-reviewed by S Madani, P Elkin; comments to author 14.10.20; revised version received 24.11.20; accepted 30.11.20; published 26.01.21.

Please cite as:
Gaudet-Blavignac C, Foufi V, Bjelogrlic M, Lovis C
Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for Processing Free Text in Health Care: Systematic Scoping Review
J Med Internet Res 2021;23(1):e24594
URL: http://www.jmir.org/2021/1/e24594/
doi:10.2196/24594
PMID:33496673

Viewpoint

Building Health Services in a Rapidly Changing Landscape: Lessons in Adaptive Leadership and Pivots in a COVID-19 Remote Monitoring Program

Celia Violet Laur[1*], MSc, PhD; Payal Agarwal[1,2*], MD; Geetha Mukerji[1,3], MD, MSc; Elaine Goulbourne[4], MSc; Hayley Baranek[1], MPH; Laura Pus[4], MBA; R Sacha Bhatia[1,5], MD, MBA; Danielle Martin[2,4,6], MD, MPP; Onil Bhattacharyya[1,2], MD, PhD

[1]Women's College Hospital Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada
[2]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[3]Department of Medicine, Division of Endocrinology & Metabolism, University of Toronto, Toronto, ON, Canada
[4]Women's College Hospital, Toronto, ON, Canada
[5]University Health Network, Toronto, ON, Canada
[6]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Onil Bhattacharyya, MD, PhD
Women's College Hospital Institute for Health System Solutions and Virtual Care
76 Grenville Street
Toronto, ON, M5S 1B2
Canada
Phone: 1 (416) 323 6400
Email: onil.bhattacharyya@wchospital.ca

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e31044

Abstract

Adaptive leadership has become an essential skill for leaders in health systems to respond to the COVID-19 pandemic as new knowledge emerges and case counts rise, fall, and rise again. This leadership approach has been described as an iterative process of taking a wide view of the situation, interpreting the meaning of incoming data from multiple directions, and taking real-time action. This process is also common in start-ups, which attempt to create new products or services of uncertain value for consumer markets that may not yet exist. Start-ups manage uncertainty through "pivots," which can include changes in the target group, need, features, or intended benefit of a product or service. Pivots are large changes that account for the high likelihood of getting something wrong during development, and they are distinct from the "tweaks" or small tests of change that define quality improvement methodology. This case study describes three pivots in the launch of a remote monitoring program for COVID-19. Adaptive leadership helped inform strategic decisions, with pivots providing a framework for internal and external stakeholders to articulate options for changes to address shifting needs. There is considerable uncertainty in the appropriate design and implementation of health services, and although this case example focuses on the use of adaptive leadership and pivots during a pandemic, these strategies are relevant for health care leaders at any time.

(J Med Internet Res 2021;23(1):e25507) doi:10.2196/25507

KEYWORDS

adaptive leadership; pivots; acute care; COVID-19; leadership; remote monitoring; monitoring; health service; framework

Introduction

Addressing the COVID-19 pandemic has required many shifts in strategy as new knowledge about the disease, its trajectory and spread, and its treatment has emerged. Health system leaders have been required to adapt health care delivery to clinical and system uncertainty as well as to the changing demographics of the populations that are most affected by the disease. For example, in mid-March 2020, the government of the Canadian province of Ontario encouraged the transfer of stable hospitalized patients to long-term care to prepare for a surge in hospital admissions due to COVID-19 [1]. By mid-April 2020, hospitals were mostly empty, so these transfers were paused; by mid-May, no patients were being transferred from hospitals to long-term care, and 82% of the deaths due to COVID-19 in Canada were occurring in long-term care facilities [2]. Protecting the capacity of acute care hospitals to prepare for the surge in cases led to strategic shifts toward, and then away from, transfers to long-term care in reaction to new information.

Adaptive challenges, or problems that cannot be solved by applying "current technical know-how or routine behavior," can be managed using adaptive leadership, as first proposed by Heifetz in 1994 [3]. Key aspects of this leadership style include taking a wider view of the situation, interpreting what is really going on, and taking real-time action to rapidly ameliorate the situation in response to the perceived or projected needs [3]. In other words, adaptive leadership is about *anticipating* future needs, trends and options; *articulating* these needs to build collective understanding and support for action; *adapting* to allow continuous learning and the adjustment of responses as necessary; and having *accountability*, including transparency in decision-making processes and openness to challenges and feedback [4]. These skills are all essential during a pandemic [3,5-8].

The level of uncertainty in health care has risen substantially during the COVID-19 pandemic. Practical approaches to managing extreme uncertainty may come from groups that are accustomed to this management, such as the founders of start-up companies. Start-ups are highly practiced in adaptive leadership because they develop products and services that do not yet exist and may not be effective for a consumer market that may not materialize. To address this challenge, start-ups have operationalized many adaptive strategies, most notably the concept of "pivots," which are used to match a service to a public need [9]. These include large changes to programs, such as narrowing or expanding the set of functions, changing customer segments, focusing on a different customer need, or changing delivery channels [9]. For example, the makers of Burbn, an app that allowed users to check in, post their plans, and share photos, noted that use of the first two functions was limited; therefore, they "zoomed in" on the photo-sharing feature and relaunched as Instagram [10]. YouTube began as a video dating website and successfully "zoomed out" to become the video streaming service known today [11]. When applied to health care, pivots may help organizations articulate different options that can be tested quickly to meet health system needs during times of extreme clinical and system uncertainty.

Pivots and adaptive leadership appear to be complementary. Adaptive leadership supports the program team in making required changes in values, beliefs, and behaviors so they can continue to meet the needs of their patients and of the health care system. Ideally, this approach helps teams to create meaning for large changes, potentially protecting against change fatigue, and positions the management of constant change as a core skill for the teams. Pivots provide a framework to facilitate difficult strategic decisions.

In March 2020, as the number of people testing positive for COVID-19 in Canada was rising [12], Women's College Hospital, an academic ambulatory hospital in downtown Toronto, a diverse city with a population of approximately 3 million, saw a need to support people in the community who tested positive for COVID-19. This need led to the development of COVIDCare@Home (CC@H), a remote monitoring program to support people who test or are presumed to be positive for COVID-19 in their homes. CC@H offers remote monitoring using telephone and video visits 7 days a week by an interprofessional, family medicine–led team, which aimed to follow patients during the acute phase of the illness (typically 14 days from symptom onset) or until they were discharged to community-based care from their primary care provider. To accelerate the process, senior hospital leaders enabled program leads to make rapid, informed, and strategic choices, drawing on principles of user-centered design with patient and provider interview feedback collected during the initial development [13,14]. This case study describes the application of adaptive leadership and three of the key pivots that allowed CC@H to launch within one week and adapt to the changing needs of patients with COVID-19 during a period of rapid changes in clinical and health system needs. A timeline of the pivots in relation to COVID-19 case rates in Toronto is included in Figure 1. Descriptions of the named pivots are visualized in Figure 2, and detailed examples based on CC@H are provided in Table 1.

Figure 1. Mapping the pivots of the COVIDCare@Home program to an approximation of the number of cases (confirmed and probable) of COVID-19 in Toronto, Ontario, Canada, from March to October 2020.

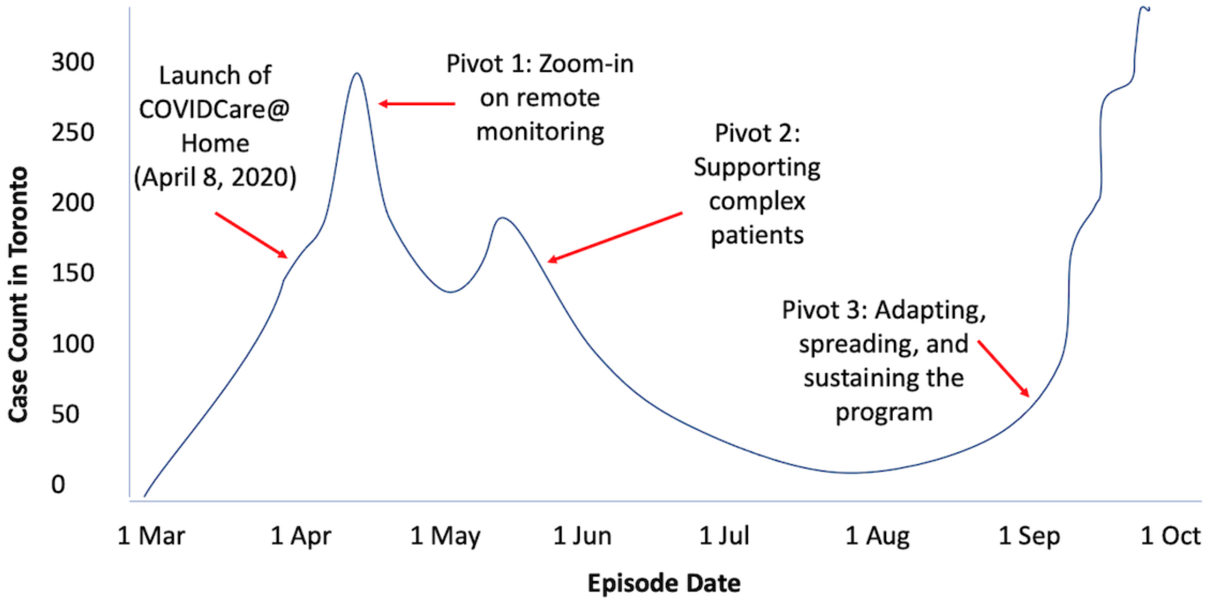


Figure 2. Names and brief descriptions of pivots. Images were obtained from the Noun Project [15].

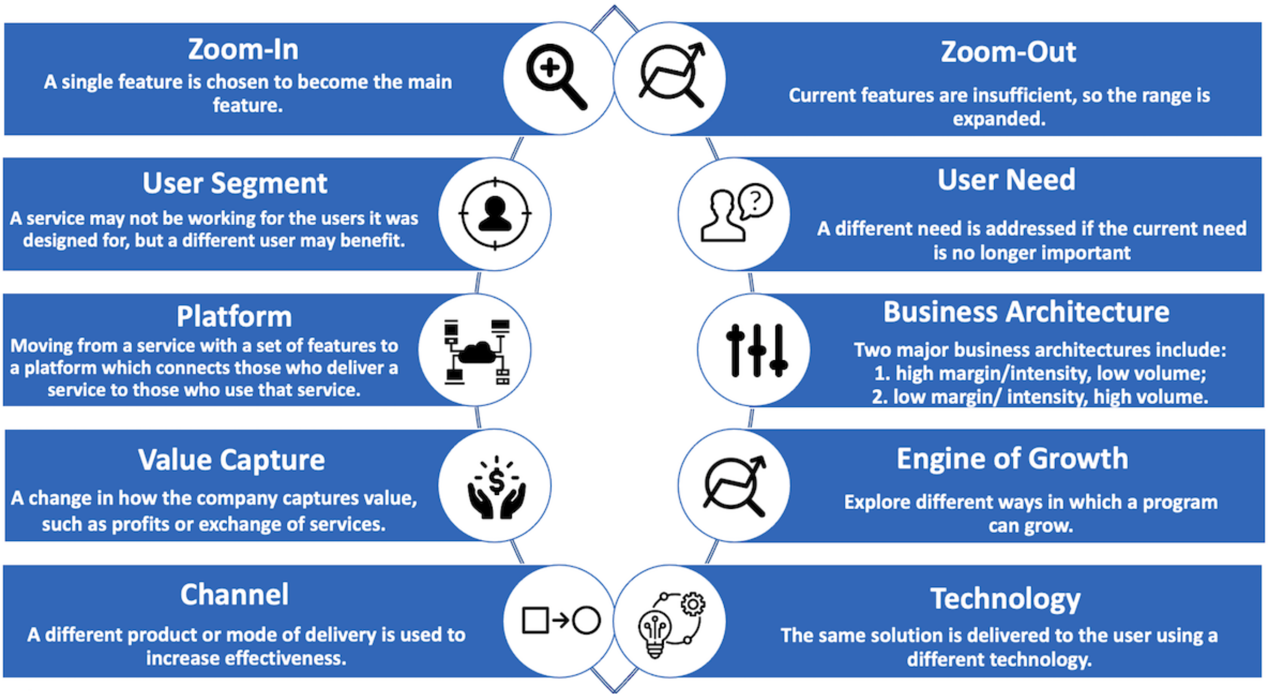


Table 1. Names, descriptions, and examples of the pivots that are mainly relevant to COVIDCare@Home. The pivots were named by Ries [9] and adapted for the health care context by the authors. "User" can refer to patients, customers, or other groups.

Pivot category and name	Description	Example
Pivots applied to CC@H		
Zoom-in	A single feature is chosen to become the main feature and everything else is cut away, thus optimizing delivery of this feature and its value proposition.	The CC@H team quickly "zoomed in" to focus on remote monitoring of community-based patients, while other strategies were deprioritized.
Business architecture	The two major business architectures include (1) high margin/intensity, low volume and (2) low margin/intensity, high volume. These approaches cannot be applied simultaneously.	The CC@H program initially provided low-intensity care to a high volume of patients because the team initially predicted high demand, then pivoted to low volume/high intensity when the COVID-19 case counts decreased and the patients were found to be more medically and socially complex.
Engine of growth	There are different ways in which a program can grow, such as changing the cost structure to make better use of existing resources; encouraging policy changes to generate new revenue sources; or developing a hub-and-spoke model to support replication in other sites.	The second wave of COVID-19 required CC@H to move from a short-term program with high resource use to a long-term sustainable program with limited resources, such as by decreasing reliance on the use of high-cost staff such as physicians.
Channel	Changing to a different product or mode of delivery to increase effectiveness.	Many health care institutions changed their primary channel of delivery during the COVID-19 pandemic when moving from in-person to virtual patient visits. Another example is the movement of CC@H from primarily using video visits with patients to using digital surveys to triage patients.
Technology	The same solution is delivered to the user using a completely different technology, such as when new technology is available at better value.	The CC@H team began sending pulse oximeters and thermometers to support monitoring rather than only using paper-based systems for tracking physiological parameters.
Pivots that *could* be applied to CC@H		
Zoom-out	The reverse of the zoom-in pivot. When the current features are insufficient for the user, the range of features is expanded.	To support the higher-intensity approach, CC@H could expand their services, such as by providing home care in addition to remote monitoring for patients at higher risk in the community.
User segment	A service may not be interesting to the users it was designed for, but early insights suggest a different user may benefit.	In the future, CC@H could change their patient population, such as to focus on postdischarge or complex patients who may receive more benefit from this approach.
User need	If early user feedback shows that the problem being solved is not very important, the team may pivot to address a different need.	As evidence of post-acute COVID-19, also known as "long COVID," emerges [16] and emphasis on new cases decreases, CC@H may move away from focusing only on supporting recently diagnosed patients to include longer follow-up of patients with continuing symptoms.
Platform	A service with a set of features changes to a platform that connects those who deliver a service to those who use that service.	CC@H could shift to a remote monitoring platform that connects specialty services to a multidisciplinary team to remotely monitor patients with different conditions.
Value capture	The is a monetization or revenue model. Leaders change how the company captures value, such as increased focus on profits or exchange of services.	CC@H could change to offering a paid service to other institutions or include it as part of inpatient care for hospital partners to allow for reimbursement in a bundled payment.

[a]CC@H: COVIDCare@Home

Pivot 1: Zooming In on Remote Monitoring

In Ontario, modeling data presented on April 3, 2020, projected that the demand for intensive care unit beds would exceed the capacity by mid-April, with a projected 80,000 cases by April 30 if current measures were followed [17]. There were concerns that hospitals would become overwhelmed. The literature contained one well-described model for remote monitoring in primary care [18]. In a desire to keep people out of hospital who could safely care for themselves at home with support, CC@H was launched by Women's College Hospital on April 8, 2020. Partners in this program included the Department of Family and Community Medicine at the University of Toronto and Mount Sinai Hospital, an acute care academic hospital that is part of the Sinai Health System.

Initial strategies of CC@H included setting up a telephone line to provide primary care providers with access to expertise in social work, pharmacy, nurse navigation, general internal medicine, respirology, and psychiatry. The team also developed resources to support early discharge from acute care and developed a protocol for remote monitoring of patients in the community [19]. CC@H quickly "zoomed in" (Table 1) on remote monitoring of community-based patients within the Greater Toronto Area who tested positive for COVID-19, while other strategies were deprioritized. This zoom-in pivot enabled the team to focus resources on refining remote monitoring processes, including video and telephone visits, methods for remote triaging, clinical pathways to address symptoms, and use of devices, such as sending pulse oximeters and thermometers to patients. Remote monitoring services followed 2020 recommendations from Greenhalgh et al [18] and were made available 24 hours per day for up to 14 days. Details are published elsewhere [20].

For initial staffing, a primary care, team-based approach was used, relying on redeployed physicians and staff from Women's College Hospital, primary care residents, and a multidisciplinary team (MDT) of providers. The MDT included nurses, a pharmacist, social workers, mental health workers, and other available specialists, who worked together to remotely address clinical needs as well as the social determinants of health of the patients.

Pivot 2: Supporting Complex Patients

The original aim of CC@H aligned with system projections that support would be needed for a high volume of patients who tested positive for COVID-19 at a low intensity, including occasional contact with mechanisms for escalation as needed [17]. However, with the initial health system focus on protecting acute care [1] and then on long-term care [2], there was delayed recognition of the need to support underserved populations who were at high risk, such as those in congregate living settings or whose low incomes, precarious work, or housing situations made their social situations particularly complex [21-23]. For CC@H, this meant that fewer patients were admitted to the program than anticipated; however, those who were admitted had complex needs beyond their COVID-19 diagnosis [20].

By early May, half of the patients admitted to CC@H had one or more comorbidity [20], and 56% belonged to occupational groups who are more likely to contract COVID-19 (such as personal support workers, shelter workers, and cleaners) and to have social challenges (such as food insecurity or lack of access to financial support), which increased the risk of poor health outcomes [20,21]. For these reasons, the program pivoted from high patient volume with low patient contact to low patient volume with high service intensity. There was an average of 4.4 visits per patient in the first month [20], with visits focused on monitoring symptoms and addressing medical and social needs, and subsequent visits were scheduled at the end of each visit based on patient preferences and clinical judgment. This type of shift in intensity is called a business architecture pivot (Table 1), which posits that a business can be low-margin, high-volume or high-margin, low-volume but not both [9]. The analogous situation in health care is a shift in the intensity and volume of a service, which in this case allowed for a more patient-centered approach that supported complex patients. The pivot facilitated use of services such as access to mental health support (ie, brief counseling and access to community resources), navigation of government support (ie, providing information on what financial or other programs the patient was eligible for and how to apply), and strategies to address food insecurity (ie, providing information on grocery delivery services, food banks, and other local initiatives available during the pandemic).

Regarding staffing changes, as the program grew, the original plan was to recruit more physicians and registered nurses (RNs) to support the high volume of patients with COVID-19–specific needs. However, with the focus on supporting patients with complex needs, a nurse practitioner (NP) was assigned instead. The NP could focus on case management specifically for complex patients who needed more intense support and delivered clinical care when the number of patients in the program increased. The emphasis was on comprehensive care, and social workers and mental health professionals also became more involved in case management.

Pivot 3: Adapting, Spreading and Sustaining the Program

When the number of COVID-19 cases decreased across Ontario in July and August 2020 [24], CC@H responded by ramping down; redeployed staff were allowed to return to their original roles, and many participating primary care physicians returned to their prepandemic practice models. The program remained nimble, retaining the ability to service higher volumes if needed. In September, when a second wave of cases began to build with predominately younger and lower-risk patients [24], a new emphasis was placed on improving triage. This approach included light-touch digital monitoring systems, such as electronic survey questions for low-risk patients, to help triage patient risk and increase the frequency of monitoring.

This second wave of COVID-19 cases necessitated an engine of growth pivot (Table 1), moving from a short-term program with high resource use and access to redeployed staff to a long-term sustainable program with limited resources and a staffing model that did not rely on redeployment. To achieve this change, the program leaders decreased reliance on the use of high-cost staff such as physicians. With increasing clinical confidence, the program team felt more comfortable with minimal physician contact for low-risk patients, enabling physicians to prioritize complex and high-risk patients. "Digital Care Coach," an electronic medical record (EMR) tool that includes a symptom questionnaire to enable low-touch remote monitoring, was prepared, and triaging guidelines (criteria for low-, medium-, and high-risk patients based on symptoms, patient history, and clinical judgment) were adapted to enable longer times between virtual visits. This EMR tool monitored symptoms of low-risk patients and provided educational materials, allowing the program to provide care to more patients while maintaining staffing levels and helping providers to focus on higher-risk patients.

Regarding staffing changes and communication strategies, to run a more sustainable program, the number of physicians involved in the program did not increase in proportion to the number of patients; instead, the program relied on more NPs and RNs to manage lower-risk patients, supported by digital tools such as Digital Care Coach. Communication strategies also became more sustainable, with the use of a patient roster and weekly clinical case conferences rather than daily huddles. The emphasis shifted to effective use of time and resources for a longer-term program.

Barriers and Enablers

Significant enablers of the rapid launch of CC@H included the suspension of many elective activities in the hospital as well as a sense of urgency. The leadership commitment to rapid action helped overcome the typical barriers and delays associated with building new programs in large organizations. Senior leaders worked closely with the program lead to ensure the necessary staffing, resources, and information technology support were available. The program lead facilitated integration of the EMR system into the program workflow, responded to stakeholder feedback in real time, developed more efficient processes for care delivery, and thus built trust among the team that enabled future pivots. Everyone involved saw the need for this program and worked through several hurdles to meet this need. Among those hurdles was adapting the EMR system to meet the changing needs of the program, which led to a steep learning curve for providers who were accustomed to a different EMR.

Maintaining the appropriate staffing levels was also challenging due to the fluctuating case numbers. A core group of physicians were involved for several months at a time, varying their number of hours per week with the program, rather than adding new providers. The social work and mental health professionals and pharmacists remained constant throughout; however, there was high turnover among RNs and NPs. Providers were flexible, moving between several programs across Women’s College Hospital based on program needs and provider skills. Providers were aware of this shifting need, and training was provided to facilitate transitions. With increasing focus on the social determinants of health, more mental health and social workers were needed, and the team was required to keep up to date on the services that were still open to know which service to recommend.

The rapid speed at which pivots occurred was both an enabler and a barrier. Changes were enacted quickly (typically within a few days) to meet the needs of patients, which limited the time for team consultation and led to a more top-down approach. The daily huddles and weekly meetings enabled the team to be informed of changes quickly and to be involved in ensuring that the changes worked for them. Although decisions to pivot were made by leadership, those decisions were informed by the team and adapted based on their continuous input in the huddles.

The decision to make this program primary care–led was another significant enabler, as it allowed for a holistic approach to care, addressing issues beyond the COVID-19 diagnosis, including social determinants of health. The primary care approach may help with the spread of the program, as it can be used in low-resource settings and in any primary care–led facility, such as long-term care. The program could also be adapted for patients who do not have COVID-19 and can be integrated into the general delivery of primary care. This remote monitoring approach did have several barriers, particularly regarding the use of technology, as many patients were unable to access video visits due to device compatibility, internet bandwidth, and other reasons. The team had to rely on different methods, including providing care via telephone or other video calling methods, such as FaceTime.

System-level enablers included increased access to human resources for pandemic-specific programs. The launch of medical billing codes for telephone and video visits facilitated physician involvement though financial renumeration of virtual care services provided. The availability of medical residents whose rotations were suspended was key, as the residents were able to serve in the model and quickly adapted to new systems and ways of working. System-level barriers included the limitations of billing codes to support case conferences and coordination.

Discussion

The rapid launch and strategic pivots enacted through adaptive leadership in CC@H enabled the team to continue to meet the needs of their patients through different waves of the pandemic. This provides a concrete example of how adaptive leadership in health care can support important outcomes in times of uncertainty [5-7].

There is considerable uncertainty in the appropriate design and implementation of health services, and although this case example focuses on the use of adaptive leadership and pivots at the organization level during a pandemic, these strategies are relevant for health care leaders at any time. These leadership skills can be learned, and the use of specified pivots may help describe options for major modifications based on emerging evidence and facilitate this adaptive approach in practice. This case focuses on individual- and organizational-level changes, and further work is needed to consider applicability at the population level, such as how elected leaders may use this approach and how to prepare the population for these rapid changes. Using this approach takes practice and courage. While the literature provides basic steps [3], there are few documented examples of the types of choices that are involved in practicing this form of leadership.

Training in the design and deployment of new services can facilitate this leadership strategy, as can the use of real-time data in implementation and outcome measures to guide decisions. Training for all staff on this adaptive approach would also be beneficial so the team members can be ready for rapid change, understand their role in the change process, and know when and how to provide feedback. This rapid feedback process will be important for staff and leadership to ensure that the pivot aligns with the needs of the team and that they have the trust of the team to go forward. The "balcony view" clearly articulated in adaptive leadership theory also acknowledges the need for system-level thinking to meet the needs of the health care system, providers, and patients. The combination of adaptive leadership and pivots provides a mechanism for making major changes in a complex system such as health care.

Additional contextual factors may have enabled the use of adaptive leadership in CC@H, such as initial close communication with senior leadership, access to staff from different disciplines, and a dedicated program lead who had training in systems engineering and had worked at start-ups. Although some of these factors are difficult to reproduce or sustain, clear communication from senior leadership and trust among team members are key and are achievable in other settings to facilitate use of adaptive leadership.

Comparison to other COVID-19 remote monitoring programs highlights directions that could have been taken by CC@H. For example, a model from Minneapolis used newsfeeds with reminders and daily check-in questionnaires about symptoms. Initially, CC@H started to develop a similar dashboard; however, while the Minneapolis program was monitoring over 1300 patients [10], CC@H only had 100 patients in the first month. Without the change in intensity (Pivot 2), CC@H would have continued to provide light-touch care for patients and would not have been able to provide the comprehensive care and case management needed for patients with complex needs at the time. However, in the second wave (Pivot 3), use of dashboards and low-touch monitoring became a priority to make the program more sustainable given the rising number of low-risk patients.

Conclusion

The CC@H program is using adaptive leadership and pivots to nimbly adapt to meet the changing needs of their patients during this time of clinical and system uncertainty, demonstrating the value of this approach. By using pivots as a framework for large strategic changes rather than small tweaks, pivots can provide direction and meaning to support health system leaders as they quickly adapt to changing needs in health care. This combination of adaptive leadership and pivots is broadly relevant at most levels of health care leadership.

Acknowledgments

The authors wish to thank everyone involved in the development and running of the COVIDCare@Home program, including providers, staff, and patients.

Authors' Contributions

CL and PA led the writing of this paper, with oversight and input from GM and OB. PA is the lead of the CC@H program. DM, SB, LP, EG, and HB were all involved in the development and/or running of CC@H and provided input into the drafting of this manuscript. All authors approved the final version.

Conflicts of Interest

None declared.

References

1. COVID-19 action plan: long-term care homes. Government of Ontario. 2020 Apr 15. URL: https://www.ontario.ca/page/covid-19-action-plan-long-term-care-homes [accessed 2020-09-09]
2. Sibbald B. What happened to the hospital patients who had "nowhere else to go"? CMAJ 2020 Jun 01;192(22):E614-E615 [FREE Full text] [doi: 10.1503/cmaj.1095873] [Medline: 32575052]
3. Heifetz R. Leadership Without Easy Answers. 1st Edition. Cambridge, MA: Belknap Press; 1994.
4. Ramalingam B, Nabarro D, Oqubuy A, Carnall D, Wild L. 5 Principles to Guide Adaptive Leadership. Harvard Business Review. 2020 Sep 11. URL: https://hbr.org/2020/09/5-principles-to-guide-adaptive-leadership [accessed 2001-01-08]
5. Ramalingam B, Wild L, Ferrari M. Adaptive leadership in the coronavirus response: bridging science, policy and practice. ODI. 2020 Apr. URL: https://www.odi.org/publications/16817-adaptive-leadership-coronavirus-response-bridging-science-policy-and-practice [accessed 2021-01-08]
6. Baker E, Irwin R, Matthews G. Thoughts on Adaptive Leadership During the COVID-19 Pandemic. JPHMP Direct. 2020 Mar 19. URL: https://jphmpdirect.com/2020/03/19/thoughts-on-adaptive-leadership-during-the-covid-19-pandemic/ [accessed 2021-01-08]
7. Warraich S, Olabi B, Azhar B, Tanzeem SF, Fischer M. Medical leadership in the NHS during the COVID-19 pandemic. Br J Hosp Med (Lond) 2020 Jun 02;81(6):1-3 [FREE Full text] [doi: 10.12968/hmed.2020.0250] [Medline: 32589538]
8. Haeusler J. Medicine needs adaptive leadership. Physician Exec 2010;36(2):12-15. [Medline: 20411840]
9. Ries E. 10 Classic Strategies For A Fast, User-Focused Company Reboot. Fast Company. 2012 Jun 04. URL: https://www.fastcompany.com/1669814/eric-ries-10-classic-strategies-for-a-fast-user-focused-company-reboot [accessed 2020-08-23]
10. Blystone D. The Story of Instagram: The Rise of the #1 Photo-Sharing Application. Investopedia. URL: https://www.investopedia.com/articles/investing/102615/story-instagram-rise-1-photo0sharing-app.asp [accessed 2020-10-07]

11. Enright K. The 4 Best Pivots in Tech History. StarterNoise. 2017 Feb 06. URL: https://www.starternoise.com/4-best-pivots-tech-history/ [accessed 2020-10-07]
12. Epidemiological summary of COVID-19 cases in Canada: Hospitalizations, intensive care unit (ICU) admissions and deaths. Government of Canada. 2020. URL: https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html#a7 [accessed 2020-10-08]
13. Chokshi SK, Mann DM. Innovating From Within: A Process Model for User-Centered Digital Development in Academic Medical Centers. JMIR Hum Factors 2018 Dec 19;5(4):e11048 [FREE Full text] [doi: 10.2196/11048] [Medline: 30567688]
14. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A Human-Centered Design Methodology to Enhance the Usability, Human Factors, and User Experience of Connected Health Systems: A Three-Phase Methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
15. The Noun Project. URL: https://thenounproject.com/ [accessed 2020-01-08]
16. Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. BMJ 2020 Aug 11;370:m3026. [doi: 10.1136/bmj.m3026] [Medline: 32784198]
17. COVID-19 Modelling Update. Ontario Health. 2020 Apr 03. URL: https://files.ontario.ca/po-covid-19-technical-briefing-en-2020-04-03.pdf [accessed 2021-01-08]
18. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
19. COVID-19 Strategy Update. World Health Organization. URL: https://www.who.int/publications/m/item/covid-19-strategy-update [accessed 2021-01-08]
20. Agarwal P, Mukerji G, Laur C, Chandra S, Pimlott N, Heisey R, et al. COVIDCare@Home: Lessons from a Family Medicine Led Remote Monitoring Program. medRxiv. Preprint posted online on July 25, 2020 . [doi: 10.1101/2020.07.20.20158386]
21. COVID-19 and the Social Determinants of Health: What Do We Know? Toronto Public Health. 2020 May 14. URL: https://www.toronto.ca/wp-content/uploads/2020/05/96e0-SDOHandCOVID19_Summary_2020May14.pdf [accessed 2021-01-08]
22. Laurencin CT, McClinton A. The COVID-19 Pandemic: a Call to Action to Identify and Address Racial and Ethnic Disparities. J Racial Ethn Health Disparities 2020 Jun 18;7(3):398-402 [FREE Full text] [doi: 10.1007/s40615-020-00756-0] [Medline: 32306369]
23. Kantamneni N. The impact of the COVID-19 pandemic on marginalized populations in the United States: A research agenda. J Vocat Behav 2020 Jun;119:103439 [FREE Full text] [doi: 10.1016/j.jvb.2020.103439] [Medline: 32390658]
24. Ontario COVID-19 Data Tool. Public Health Ontario. URL: https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance/covid-19-data-tool [accessed 2020-09-05]

Abbreviations

CC@H: COVIDCare@Home
EMR: electronic medical record
NP: nurse practitioner
RN: registered nurse

Edited by G Eysenbach; submitted 04.11.20; peer-reviewed by B Durafourt, A Wong, Z Aghaei, A Khaleghi; comments to author 26.11.20; revised version received 18.12.20; accepted 24.12.20; published 13.01.21.

Please cite as:
Laur CV, Agarwal P, Mukerji G, Goulbourne E, Baranek H, Pus L, Bhatia RS, Martin D, Bhattacharyya O
Building Health Services in a Rapidly Changing Landscape: Lessons in Adaptive Leadership and Pivots in a COVID-19 Remote Monitoring Program
J Med Internet Res 2021;23(1):e25507
URL: http://www.jmir.org/2021/1/e25507/
doi:10.2196/25507
PMID:33417588

Viewpoint

Going From Zero to 100 in Remote Dementia Research: A Practical Guide

Megan E O'Connell[1], PhD; Shirin Vellani[2], MN-NP; Sheryl Robertson[2], BScN, BSc; Hannah M O'Rourke[3], PhD, RN; Kathy S McGilton[2], RN, PhD, FAAN, FCAHS

[1]University of Saskatchewan, Department of Psychology, Saskatoon, SK, Canada
[2]University Health Network, Toronto Rehabilitation Institute, University of Toronto, Toronto, ON, Canada
[3]University of Alberta, Faculty of Nursing, Edmonton, AB, Canada

Corresponding Author:
Megan E O'Connell, PhD
University of Saskatchewan
Department of Psychology
9 Campus Drive
Saskatoon, SK, S7N 5A5
Canada
Phone: 1 3069662496
Email: megan.oconnell@usask.ca

Abstract

Remote approaches for dementia research are required in the era of COVID-19, but moving a research program from in person to remote involves additional considerations. We recommend using outcome measures that have psychometric properties for remote delivery, and we recommend against adapting in-person scales for remote delivery without evidence for psychometric equivalency. We suggest remote research designs that maximize benefit for participants, which could have implications for control groups. Researchers should plan for flexibility in their methods for remote research and must not assume all participants will be able to videoconference; telephone-only research is possible. We recommend performing an assessment of information communication technology infrastructure and prior exposure to this technology with each participant before making a final choice on remote methods for research. In general, researchers should adapt their methods for remote research to each participant rather than requesting participants to adapt to the researchers. Screening for sensory loss should be conducted, and the impact of this on the use of technology for remote research should be considered. In this viewpoint, we detail how individualized training is required prior to engaging in remote research, how training plans interact with cognitive impairments and, finally, the steps involved in facilitating technology-based remote data collection.

(J Med Internet Res 2021;23(1):e24098) doi:10.2196/24098

KEYWORDS

COVID-19; telehealth; videoconferencing; dementia; information communications technology

Introduction

Remote approaches for dementia research overcome barriers to participation in the era of COVID-19 requiring social distancing measures, but are also required to mitigate other factors, such as geographic barriers experienced by rural families [1-3]. In the midst of the COVID-19 pandemic, however, there are additional considerations: is this research necessary at time? Does participation involve undue stress or increase risk exposure? Does the use of remote methods undermine the quality of the research? For example, many of us are conducting remote dementia assessments, but would this method of diagnosis meet the research standards for a gold standard in validation studies? This decision-making process could have implications for research design—for example, does the use of a control group make some research untenable at this particular time given the impacts on the risk benefit analysis? Clearly, the risk benefit analysis depends on circumstances. We argue that in-person contact in the era of COVID-19 should be minimized for research, and that virtual or remote methods, which are in many situations the only option, are ethically preferred. Where an ethical decision-making process has determined the benefits to outweigh the (ideally minimal) risks, we aim in this viewpoint to provide guidance on how to move your research from in-person to remote work.

Remote Dementia Research: Should You Do It?

Remote research can refer to research conducted solely by telephone (landline or smartphone interface) or by videoconferencing. Videoconferencing can occur via telehealth networks provided by local health care (ie, videoconferencing equipment in hospital or clinic settings) or with internet-based software platforms for participants with adequate in-home information communications technology (ICT) such as Microsoft Teams, Webex, Zoom, or Skype. The steps to consider in remote research are summarized below. We draw your attention to the ordering of some steps, which might be counterintuitive. Most notably, for those using videoconferencing for remote research, we recommend choosing a videoconferencing platform after you have conducted an ICT assessment with each participant (process described in Table 1).

Table 1. Zero to 100 in remote dementia research overview.

Issue	Considerations	Recommendation
1. Theoretical assessment of the feasibility of remote measurement	Are your research outcomes adaptable for remote research without compromising evidence for validity?	Choose measures with evidence for remote validity or telephone validity or choose scales with very few adaptations from in-person administration.
2. Assessment of the risks and benefits of the research design for participants	Participants may be under additional pressure and reluctant to participate if it is not clear how they will benefit. For example, is it essential to have a control group in your behavioral intervention study?	Consider participant burden as foremost to being "pandemic-friendly." Choose a single-group repeated-measure design where individuals act as their own controls when it is unethical or not practical to have a control group, or consider use of a wait-list control group. Streamline your approach to measurement to reduce response burden.
2. Theoretical assessment of the likely ICT[a] infrastructure of your research population	Is videoconferencing possible or is telephone-based contact most likely for remote research?	Plan for flexibility in remote research—even if you prefer videoconferencing, always include telephone-only contact as a backup plan.
3. Assessment of ICT infrastructure for each participant before research participation starts	Telephones are common, but do they have a computer, smartphone, or tablet? Do they have broadband access? Do they have speakers or headphones? Do they have a microphone?	Use a screening question like "do you have a computer, tablet, or smartphone that you use to connect with others?" A consideration: budget to send necessary ICT equipment to remote participants if appropriate.
4. Consider the needs of participants with cognitive or sensory impairments, or both	We detail special considerations for sensory and cognitive impairments, but these are highly individualized to each participant.	Screen for cognitive and sensory impairments and adapt your method of remote research accordingly.
5. Consider platform for videoconferencing research	Only some videoconferencing platforms are private and secure, which is necessary to meet REB[b] approval. Consider what you need in terms of number of people joining and consider your participants' experiences.	Adapt to your participant's preferences and prior experiences with software; do not make them adapt to you.
6. Train participants for remote research	Training and support for remote research is likely required, and we detail some strategies to help with training.	Plan to spend a sizeable amount of time training participants to use new ICT equipment/platforms.
7. Obtain remote consent	We detail issues in obtaining consent, including obtaining proxy consent remotely.	The method used for remote participation should be the same as that used for informed research consent—for telephone contact, telephone consent; for videoconferenced contact, videoconferenced consent.
8. Set the scene	We detail steps required to minimize distractions during the remote visit.	Plan for communication failures by obtaining multiple methods for communication.

[a] ICT: information communications technology.

[b] REB: research ethics board.

Remote Measurement: Can You Do It?

If you use any standardized or quantified scales for research, you must first decide if you can translate these to remote administration without invalidating the measurement properties of the scale. It is possible you will need to consider using alternative assessment methods, and it might involve changing your research design to qualitative vs quantitative. Scales that have evidence for psychometric properties under the conditions of remote administration are the best choice. It is not advisable to use a scale that has established psychometric properties only for use in person and modify this for remote administration under the assumption that it is equivalent to in-person administration. We do not recommend this practice because it introduces unknown sources of measurement error or could change the validity of measurement [4,5]. Does this mean no modified tests can be used? No, but we recommend examining the literature for evidence of impact of changing mode of administration to remote methods on measurement. We recommend following the helpful guidelines suggested by a task force on good research practices for modifying patient-reported outcomes for electronic administration [5] because they suggest levels of evidence for measurement should vary based on degree of scale modifications needed for remote administration.

Notably, Coons et al [5] recommend use of prior research to determine the levels of modifications needed to take a test from being administered in person to being remotely administered. In the event that modifications to a quantified scale for remote delivery are minor (eg, response options remain the same, but a mouse click is exchanged for circling to indicate a response), Coons et al [5] recommend cognitive debriefing. Cognitive debriefing is defined as a qualitative evaluation of how items were approached to determine if the items were understood in the way the researcher or clinician intended (ie, evidence for content validity). At the other extreme, Coons et al [5] suggest that if modifications to a test are major, which could include item wording changes or item response option changes, or a psychomotor output becomes a verbal output, it is a new test and full psychometric testing is required. Moderate modifications include wording changes and changing a non–psychomotor-based visual output to a verbal one. If modifications are moderate, this could change an item's meaning or general content, and Coons et al [5] recommend equivalence testing. Equivalence testing can be done between and within subjects (eg, randomized cross-over designs), within subjects (eg, repeat administration in person and remote), or between subjects (eg, different groups who received the test in person vs by remote methods using differential item functioning methods or multigroup confirmatory factor analysis for measurement invariance of latent factors). If prior data for equivalency testing are not available, it might be impossible to gather these during the pandemic, which could impact your choice of measures for remote research. Irrespective of the degree of modification, normative comparison standards created for an in-person version of a measure cannot be applied to the remote-delivered version of a measure, even when the modifications for remote administration appear minor (O'Connell et al, unpublished data, 2021). Although this finding has more clinical than research implications, it impacts use of clinical scales for remote research. Ensuring the measures are reliable and valid for remote testing will involve additional pilot work prior to proceeding to the study at hand. At the very least, using scales which have not been tested remotely will need to be noted as a limitation to the findings of the study, which could jeopardize the study conclusions and undermine your remote research.

Remote Study Designs: Are Your Modifications "Pandemic-Friendly"?

Multiple factors may complicate evaluation of interventions, particularly in situations where a randomized controlled trial is impractical, culturally or clinically unacceptable, or ethically questionable [6]. A single-group repeated-measure design can be utilized in this situation [7], which may be the case during COVID-19. In this way, the intervention is provided to all eligible participants, and the outcomes are assessed repeatedly before (thus reflecting the no-treatment condition) and after (thus reflecting the intervention condition) the intervention delivery. While there are disadvantages to this design, researchers often forget the advantages, which may be particularly relevant in a situation where participant burden is of utmost concern, such as during the current pandemic. Care home environments, for example, have been greatly impacted by COVID-19, but some research may be absolutely critical to inform how we respond to infectious disease outbreaks in ways that both saves lives and maintains quality of life. The main advantage of the single-group design is that participants serve as their own control, a situation that (1) is consistent with the counterfactual posited as ideal for determining the causal effects of an intervention [6], (2) reduces the potential of confounding as the same participants with the same personal and health or clinical characteristics are exposed to both the control and the intervention conditions [8], and (3) decreases the number of participants needed to detect significant intervention effects in the single-group repeated-measure design [9]. The number of participants needed is reduced because multiple measurements on each participant produces more data to support inferences about change. Specifically, repeated measurement of each subject (ie, individual participant data) provide enough data to adjust for baseline imbalance between treatment and control, to account for interactions among covariates, and to account for correlations between baseline and follow-up measurements of the outcome. Moreover, utilizing design approaches that do not include control groups coincide with values of inclusion and sharing of opportunity that are required as conditions for patient and community engagement [6]. Another alternative is the wait-list control group design. With a wait-list control group design, advantages include that random assignment could be maintained and that all participants would eventually receive the intervention. A disadvantage is that the data collection period is extended (to allow for outcome data collection from both intervention and control groups), and the control group may have to wait a significant amount of time to receive a potentially beneficial intervention. This may not be deemed ethical in the urgent context of COVID-19 nor be considered "pandemic-friendly." Regardless of the design selected, the measurement approaches must be streamlined so that no participant is being asked to commit more time or energy than what is absolutely necessary to generate valid and useful knowledge that can be used to inform our response to the pandemic and beyond.

Remote Participant Contact: Can You Do It?

This step is theoretical and will be refined in later steps, but before you engage in participant contact, consider your research participants and the likelihood of their access to the ICT infrastructure. List what ICT equipment is needed for your method of remote research and consider your participant population. Are they likely to have the ICT infrastructure? Are they likely have the knowledge to use this infrastructure without inducing undue stress? The ICT infrastructure recommended for videoconferencing includes 1024 kbps bandwidth for videoconferencing, as well as newer models of smartphones, tablets, or computers with webcam; speakers or headphones; and microphones. In our experience, this ICT infrastructure is not ubiquitous in the homes of many older adults, underscoring the lack of necessary ICT infrastructure for this population during the pandemic. Telephones are almost ubiquitous and, therefore, are a low burden method of communication. However, even their presence should not be assumed (eg, many residents of care homes do not have regular access to telephones). Our prior work detailed travel burden experienced by rural families in accessing telehealth videoconferencing [2], so in our current remote research we use telephone for intake procedures. Finally, are your planned participants likely to have sensory or cognitive impairments that make interaction with remote methods more challenging, or even impossible (we discuss this in a special section below)?

Before You Engage in Remote Research, Assess the ICT Readiness of Each Participant

We propose initial participant contact should occur using the participant's preferred method of contact, which for many is the telephone. Telephones are almost ubiquitous and are, therefore, a low burden method of communication (ie, they are accessible and easy to use). Before you engage in remote research, you must be in contact with each participant to assess their appropriateness/suitability for remote contact. Videoconferencing for remote research is the closest analogue to in person and, therefore, has numerous advantages [10]. Some nonverbal cues are available, visual mouth cues can help with those who have hearing loss, and rapport can be easily established when used for dementia care [1]. We discovered, however, that videoconferencing misses many nonverbal cues [2], and we recommend the researcher aims to be extremely explicit and clear in communication and be prepared to ask often for clarification of facial expressions or subtle signs of discomfort.

The goal of your ICT readiness assessment to be conducted with each participant prior to engaging in the research process is to determine if they have the necessary ICT infrastructure. Answering this deceptively simple question is made more difficult due to the bidirectional relation between ICT and participants' comfort with and exposure to said ICT. We developed a rural technology acceptance model [3] and an Indigenous adaptation to this technology acceptance model [11], which underscore the multitude of reasons people might avoid new ICT, which includes longstanding infrastructure access barriers. If the answer to any of the following questions is no, you cannot use videoconferencing with this participant and need to consider an alternative plan such as use of the ubiquitous telephone. Recommended questions ascertain each participant's ICT access and experience with this ICT. In our work, we see many people who would not know how to answer the question: do you have highspeed internet access? We have *not* encountered anyone able to tell us if their internet exceeds the recommended 1024 kbps bandwidth for videoconferencing. The researcher may not be able to identify or control for bandwidth issues available for participants at home; however, if the researcher has access to a bandwidth of 1024 kbps, it will ensure an acceptable quality for up to 3 connections (researcher and 2 participants) [10]. A screening question we have found useful is: do you have a computer, tablet, or smartphone that you use to connect with others? If yes, do you connect with others using video? If yes, does this video freeze and make it hard to communicate? From these questions, one can get an idea of the adequacy of their ICT infrastructure and their prior exposure in day-to-day activities. The time taken to assess the ICT infrastructure can be somewhat lengthy, but it is important to address questions to help researchers better prepare themselves and the participants. For example, this ICT infrastructure assessment may suggest the need to create a step-by-step guide, the need for specific equipment (headphones and/or webcams), the need for an informal support person to help troubleshoot with the participants, or the need to use the telephone.

Special Consideration for Participants With Sensory Impairments

Few method exist to remotely assess sensory impairments, but a telephone-based hearing test service is available (ie, the National Hearing Test [12]) [13]. Sensory impairments could contraindicate remote methods, but this is highly individualized. For those with some hearing deficit, headphones may enhance communication since they can help with the amplification of the researcher's voice, which is not available in face-to-face settings [10]. Severe auditory impairments can be mitigated by using closed captioning during videoconferencing, and automated and real-time closed captioning methods are available to use with some videoconferencing platforms (eg, Zoom). One will need to manipulate the videoconferencing camera placement to ensure adequate exposure to the researchers'/clinicians' mouth-based cues, which can also mitigate hearing loss, provided, of course, the visual and auditory feed are synchronous. Headphones can help a lot with minimizing extra noise and focusing the sound, but in our experience, few participants have these at home and researchers might need to be prepared to supply these. The researcher should, however, use a system with a high-quality microphone—either a stand-alone USB microphone (situated nearer to your mouth than what would be standard on desktop computers) or a good-quality microphone on a headset. Removal of distractions can help (discussed later), but sensory challenges do not interact well with low-quality connections. Hence, we recommend considering an alternate method of remote communication if the connection is slow. Alternative methods include use of the telephone for remote research or asynchronous methods (eg, email, postal mail). Use of screens to display questionnaires can help when videoconferencing for remote research [10]; closed captioning could also be helpful for mitigating challenges in communication due to hearing loss. Additionally, we have found that visually providing a diagram with the response options to the survey questions when verbally reading them to the participant helps those with sensory or cognitive issues complete the surveys more efficiently.

If You Choose Videoconferencing for Remote Research, Which Platform Do You Use?

We recommend that the participant should not have to adapt to you; you should adapt to them for remote research. This helps to mitigate anxiety in participants and maximize the probability of success for the remote interaction, and leverages prior learning for those with memory impairments. A final reason for using a platform your participant is familiar with is to minimize the amount of new learning required because you will need to plan to train each participant (and potentially a family member or friend who can support them) in the use of the technology platform required for your remote contact.

It might, however, not be possible to solely let your participants' prior experiences and preferences guide your choice of software platform for videoconferencing. Foremost, the remote method platform needs to be private and secure, and if this is not possible, informed consent needs to address the potential loss of privacy or security. Local research ethics boards need to decide their comfort with nonsecure platforms for research. Many videoconferencing platforms are compatible with the Health Information Protection Act, the Personal Information Protection and Electronic Documents Act, and the Personal Health Information Protection Act, such as WebEx, Zoom Healthcare (note: not the open free version), NousTalk, Doxy.me, Microsoft Teams, and Pexip. Many of these platforms allow for group videoconferencing, which can allow a person living with dementia to join with a caregiver who may live separately. In addition, group-enabled platforms allow for live supervision of trainees.

Remote Contact: How to Do It

Ethical Issues

Although you might need to include the possibility of technical challenges causing frustration in your consent process, many ethical issues with remote research remain the same as for in-person research. In contrast, obtaining proxy consent when it is required in cases with diminished capacity could become more complicated under conditions of remote work. Typically for our clinical research, in-person visits are attended by someone in a position to provide proxy consent, and, as a matter of routine, we obtain assent and proxy consent when asking consent to speak with both a patient and a collateral informant or caregiver. In our diagnostic research clinic, we speak with the patient and caregiver together and separately, and we do this for our remote clinical research as well. This, however, takes multiple remote contacts sometimes spanning several days. We have not yet encountered a scenario where a live-in caregiver, such as a spousal caregiver, wishes to say something frank and their care partner will not allow them to be left to speak with us in private, but this scenario is conceivable. Attempts to time the caregiver interview during times the person living with dementia is occupied with another task might be the only method to mitigate this issue. It is also possible that a private conversation is not going to be possible, and alternate methods for private communication (ie, written) should be offered.

Despite the platform used for remote contact, you are not able to control who your participants have in their room. This has implications for privacy and confidentiality if headphones are not being used for videoconferencing or if speakerphones are used for telephone contact. We recommend a practice of introducing everyone in the room even if they are not going to be on screen or on the call.

Finally, we suggest that consent procedures should use the same procedures used for in-person contact. For in-person contact, written consent is easy and does not add to participant burden. If initial contact is over the telephone, verbal consent including its limits of confidentiality should be completed over the telephone, which can be recorded if necessary. Salmons [14] suggests that expecting the participants to download, sign, and return the consent form to the research team may "be unrealistic." Many of the people we work with do not have computer or email access precluding the opportunity to email them a consent form, and needing to sign and return a self-addressed envelope containing a consent form would create an additional burden, never mind potentially put persons at unnecessary risk in the era of COVID-19. It is possible that research ethics boards will require a hard copy of the consent be mailed to participants so they can follow it along on the day of the consent meeting, or may require having a witness present with the participant whose name is recorded on the consent form, and then allow the consent conversation to be digitally recorded with the final step of a researcher-signed copy sent to the participants via post. This could introduce undue burden and restrict participation by some. We hope all research ethics boards consider the barriers to participation in remote service delivery and research that can be created by requiring written consent or paper-based consent processes, and the implications this has for those most vulnerable to COVID-19 and for those with limited ICT access. The research ethics board at the University of Saskatchewan, for example, has verbal consent procedures for research, which facilitates engagement in research by rural families with few ICT resources.

Training and Troubleshooting

If you choose a method for remote contact that is not familiar to the participant or if you are asking them to use this method for remote contact in ways that are novel to them (ie, answering Likert-style questions), be prepared to spend time training them before you begin the remote research. We have detailed how independent use of videoconferencing can occur for the technologically inexperienced with remote training support [2]. We recently moved to an internet-based videoconferencing platform to engage in a socialization to mitigate isolation in the era of COVID-19, and we are prepared with training videos and solutions to common connection problems, and can screenshare to show participants how to interact with the new platform. Unfortunately, we have discovered that screenshare only works well if we are all using the same interface (Windows or Mac), computer, tablet, or smartphone. We recommend sending a step-by-step guide or having screenshots of common problems sourced from a multitude of helpful YouTube videos that you can share on a moment's notice (a second piece of technology to source this troubleshooting information can be helpful). Regardless of your method for remote communication, we recommend planning for the worst, which could include a catastrophic failure in technology. We also advise you have a telephone contact number as a back-up and communicate how you will use this during the event of a disconnection.

It might be prudent to screen participants for cognitive impairments; numerous methods for screening for cognitive impairments have been validated for remote delivery [15,16]. Neuropsychological deficits can interact with technology use and learning of new technology; consequently, we recommend leveraging ICT methods with which the participants have prior exposure if there are cognitive impairments [17]. Persons living alone with cognitive impairment (eg, mild cognitive impairment or mild to moderate dementia or other etiologies) will require several training support sessions to engage in videoconferencing using a platform that is new to them. It is possible that persons living with moderate cognitive impairment will need remote training from an expert in cognitive rehabilitation. Common techniques from cognitive rehabilitation can train people to use new technology, even if they have marked anterograde amnesia [17], and cognitive rehabilitation can be delivered remotely to persons living with mild cognitive impairment or mild to moderate dementia [18,19]. If cognitive rehabilitation is needed to train persons living alone with mild cognitive impairment or dementia to use videoconferencing and this intervention support is not available, we recommend engaging the telehealth suites if they are available through local health care agencies or use the telephone for your remote research. Use of additional volunteers within a participant's existing "pandemic bubble" to act as technology support can be useful and safe. For example, we spoke over the phone with a home care aid who was on site and helped with some technological challenges to facilitate the remote training needed for a participant to engage in remote research.

Are You Ready to Go? Set the Scene

Prior to your first remote research session, it is critical that you discuss setting the scene with your participant. This brief step assumes you have had some training and practice sessions with your participant ahead of the first data collection visit if it is over videoconferencing, and you have ensured the camera angle, volume, and microphone placement are ideally set up for your encounter. You and your participant should agree on a remote contact time that is likely to be distraction-free (eg, is not during a favorite television show) and expectations regarding multitasking need to be explicit—state, "if something comes up that you have to deal with let me know and we can reschedule." In our experience, people are less likely to multitask during videoconferencing, but we need to be more explicit in our expectations for the remote contact over the telephone. Nevertheless, telephone calls can interrupt participants during videoconferencing. Asking participants to turn off the ringer can complicate your back-up plan to use the telephone in the event of a failure in technology. We ask our participants to turn off ringers but agree on a plan to turn them back on in the event that the videoconferencing fails.

In the era of COVID-19, many clinicians and researchers are conducting remote communication from within their own homes, providing yet another venue for distractions and yet another limit of confidentiality if working from multiperson households. Examine the view your participant has of you, simplify your backdrop, and maintain a self-view lest a wayward pet comes into view and provides welcomed relief, a distraction, or both.

Advantages Afforded by Remote Research

Remote methods for research offer several advantages beyond meeting physical distancing requirements and traversing all geographical boundaries, which reduces travel burden for participants. Foremost, research diaries of the adaptation for remote research could, in themselves, be a research output. Remote methods allow for alternative methods of participation, for example, access to language interpreters and translators could be facilitated within the remote framework. Audio and audio-video recording is seamless with many remote methods to facilitate transcription for analysis, training of staff, and supervision of staff, provided, of course, this is conducted within a secure framework and is consistent with behavioral ethics protocols and consent processes. Remote therapy will also allow for a more careful attention to treatment fidelity, which is a key methodological requirement of any sound intervention and can enhance internal validity of the trial. In a review of treatment fidelity in nonremote behavioral intervention studies, whether or not an intervention was delivered was often reported, but little attention to the other elements of fidelity such as training of staff and whether the participant actually received the intervention as intended were lacking [20]. These authors suggest that inattention to treatment fidelity may be due in part to the additional resources required to assess treatment fidelity; live supervision could mitigate some of these concerns. Live supervision conducted remotely could occur via the researcher joining a videoconference call, for example, to assess fidelity of delivery by a research assistant.

We hope to have conveyed that moving from in-person to remote dementia research is time consuming and must be completed with careful consideration, but is a worthwhile endeavor.

Acknowledgments

MEO conceived the idea with KSM, but all authors contributed to the further genesis and organization of the manuscript over a series of team meetings. SR, SV, and MEO conducted the literature review. MEO wrote the first draft of the manuscript, and all authors revised the manuscript.

Funding was provided by the Canadian Consortium on Neurodegeneration in Aging (CCNA). CCNA is supported by a grant from the Canadian Institutes of Health Research with funding from several partners including the Saskatchewan Health Research Foundation, the Centre for Aging and Brain Health, and the Alzheimer Society of Canada (ASC). The ASC COVID Task Force (MEO and KSM are members) and their support for CCNA Social Inclusion/Stigma theme work in CCNA was a particular catalyst for this work.

Conflicts of Interest

None declared.

References

1. Morgan DG, Crossley M, Kirk A, McBain L, Stewart NJ, D'Arcy C, et al. Evaluation of Telehealth for Preclinic Assessment and Follow-Up in an Interprofessional Rural and Remote Memory Clinic. J Appl Gerontol 2011 Jun;30(3):304-331 [FREE Full text] [doi: 10.1177/0733464810366564] [Medline: 24966449]
2. O'Connell ME, Crossley M, Cammer A, Morgan D, Allingham W, Cheavins B, et al. Development and evaluation of a telehealth videoconferenced support group for rural spouses of individuals diagnosed with atypical early-onset dementias. Dementia (London) 2014 May;13(3):382-395. [doi: 10.1177/1471301212474143] [Medline: 24339063]
3. O’Connell M, Scerbe A, Wiley K, Gould B, Carter J, Bourassa C, et al. Anticipated needs and worries about maintaining independence of rural/remote older adults: Opportunities for technology development in the context of the double digital divide. Gerontechnology 2018 Sep 30;17(3):126-138. [doi: 10.4017/gt.2018.17.3.001.00]
4. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Arch Clin Neuropsychol 2012 May;27(3):362-373 [FREE Full text] [doi: 10.1093/arclin/acs027] [Medline: 22382386]
5. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
6. Henry D, Tolan P, Gorman-Smith D, Schoeny M. Alternatives to Randomized Control Trial Designs for Community-Based Prevention Evaluation. Prev Sci 2017 Aug;18(6):671-680. [doi: 10.1007/s11121-016-0706-8] [Medline: 27600286]
7. Ibrahim S, Sidani S. Fidelity of Intervention Implementation: A Review of Instruments. Health 2015;07(12):1687-1695. [doi: 10.4236/health.2015.712183] [Medline: 28213585]
8. Lawlor DA, Smith G, Bruckdorfer K, Kundu D, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? The Lancet 2004 May 22;363(9422):1724-1727. [doi: 10.1016/S0140-6736(04)16260-0] [Medline: 15158637]
9. Riley RD, Kauser I, Bland M, Thijs L, Staessen JA, Wang J, et al. Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. Stat Med 2013 Jul 20;32(16):2747-2766. [doi: 10.1002/sim.5726] [Medline: 23303608]
10. Tuerk P, Shore P, Ronzio J. Technologies and Clinical Videoconferencing Infrastructures: A Guide to Selecting Appropriate Systems. In: Tuerk P, Shore P, editors. Clinical Videoconferencing in Telehealth. Behavioral Telehealth. Cham: Springer; 2015.
11. Starblanket D, O’Connell M, Gould B, Jardine M, Ferguson M, Bourassa C. Defining technology user needs of Indigenous older adults requiring dementia care. Gerontechnology 2019 Sep 28;18(3):142-155. [doi: 10.4017/gt.2019.18.3.002.00]
12. National Hearing Test. URL: https://www.nationalhearingtest.org/wordpress/?page_id=2730 [accessed 2021-01-22]
13. Watson CS, Kidd GR, Miller JD, Smits C, Humes LE. Telephone screening tests for functionally impaired hearing: current use in seven countries and development of a US version. J Am Acad Audiol 2012;23(10):757-767. [doi: 10.3766/jaaa.23.10.2] [Medline: 23169193]
14. Salmons J. Designing and conducting research with online interviews. In: Cases in Online Interview Research. Thousand Oaks, CA: SAGE Publications; 2012:1-30.
15. Clark K, St John P. Virtual approaches to cognitive screening during pandemics. JGS Journal of CME 2020;10(1) [FREE Full text]
16. Frank C, St John P, Molnar F. Screening tools for virtual assessment of cognition. Can Fam Physician 2020 Jul;66(7):502-503 [FREE Full text] [Medline: 32675095]
17. O'Connell ME, Mateer CA, Kerns KA. Prosthetic systems for addressing problems with initiation: guidelines for selection, training, and measuring efficacy. NeuroRehabilitation 2003;18(1):9-20. [Medline: 12719617]
18. Cotelli M, Manenti R, Brambilla M, Gobbi E, Ferrari C, Binetti G, et al. Cognitive telerehabilitation in mild cognitive impairment, Alzheimer's disease and frontotemporal dementia: A systematic review. J Telemed Telecare 2019 Feb;25(2):67-79. [doi: 10.1177/1357633X17740390] [Medline: 29117794]

19. Burton RL, O'Connell ME. Telehealth Rehabilitation for Cognitive Impairment: Randomized Controlled Feasibility Trial. JMIR Res Protoc 2018 Feb 08;7(2):e43 [FREE Full text] [doi: 10.2196/resprot.9420] [Medline: 29422453]
20. O’Shea O, McCormick R, Bradley J, O’Neill B. Fidelity review: a scoping review of the methods used to evaluate treatment fidelity in behavioural change interventions. Physical Therapy Reviews 2016 Dec 20;21(3-6):207-214 [FREE Full text] [doi: 10.1080/10833196.2016.1261237]

Abbreviations

ICT: information communications technology

Edited by G Eysenbach; submitted 03.09.20; peer-reviewed by A Hall, W Zhang; comments to author 28.09.20; accepted 18.01.21; published 27.01.21.

Please cite as:
O'Connell ME, Vellani S, Robertson S, O'Rourke HM, McGilton KS
Going From Zero to 100 in Remote Dementia Research: A Practical Guide
J Med Internet Res 2021;23(1):e24098
URL: http://www.jmir.org/2021/1/e24098/
doi:10.2196/24098
PMID:33468448

Original Paper

Digital Health Needs of Women With Postpartum Depression: Focus Group Study

Madison E Lackie[1,2], MSc; Julia S Parrilla[3], MSc; Brynn M Lavery[2], BA; Andrea L Kennedy[4,5], MD, FRCPC; Deirdre Ryan[4,5], MD, FRCPC; Barbara Shulman[4,5], MD, FRCPC; Lori A Brotto[1,2], PhD, RPsych

[1]Women's Health Research Institute, BC Women's Hospital + Health Centre, Vancouver, BC, Canada

[2]Department of Obstetrics & Gynaecology, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[3]School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[4]Reproductive Mental Health Program, BC Women's Hospital + Health Centre, Vancouver, BC, Canada

[5]Department of Psychiatry, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Lori A Brotto, PhD, RPsych
Women's Health Research Institute
BC Women's Hospital + Health Centre
Box 42B
H214 - 4500 Oak St.
Vancouver, BC, V6H 3N1
Canada
Phone: 1 604 875 3459
Email: lori.brotto@vch.ca

Abstract

Background: Although approximately 10% of new mothers in Canada develop postpartum depression (PPD), they face many barriers when accessing care. eHealth offers a unique opportunity to provide psychosocial skills and support to new mothers; however, patient populations are not consistently engaged in eHealth development processes. Thus, the diversity of women’s backgrounds and needs are often not reflected in existing tools.

Objective: This study aims to engage women from a variety of backgrounds and locations around British Columbia (BC) who have previously experienced PPD to determine the unmet psychoeducational needs of women with PPD and how a web-enabled platform used to deliver psychosocial skills and education to assist in the management of PPD could fulfill those needs.

Methods: Focus groups were conducted in 7 cities across BC with a total of 31 women (mean age 34.5 years, SD 4.9), with each group ranging from 2-7 participants. Focus groups were cofacilitated by the study coordinator and a local service provider in each community using a semistructured guide to discuss participants’ needs, ideas, and opinions as they relate to the use of technology in PPD management. Transcripts were approached inductively using thematic analysis to identify themes and qualitative description to frame what was observed in the data.

Results: A total of 5 themes were identified: bridging gaps to meet needs; providing validation to combat stigma; nurturing capacity to cope, manage, and/or reach wellness; empowering people to take ownership over their mental health; and offering customization to ensure relevance. Each theme identified a need (eg, combatting stigma) and a way to address that need using a web-enabled intervention (eg, providing validation). At the intersection of these themes was the overarching value of promoting agency for women experiencing PPD.

Conclusions: Ultimately, new mothers require accessible mental health care that promotes their agency in mental health care decision making. Our participants believed that a web-enabled intervention could help meet this need. These data will be used to guide the design of such an intervention, with the eventual implementation of this resource as a first-line management option for PPD.

(J Med Internet Res 2021;23(1):e18934) doi:10.2196/18934

KEYWORDS

postpartum depression; perinatal mental health; eHealth; women’s health; reproductive health; maternal health; qualitative research; focus groups; user-centered design; knowledge translation; self-management

Introduction

Postpartum Depression

Postpartum depression (PPD) is a leading cause of maternal morbidity and mortality in high-income countries [1] and indiscriminately affects new mothers around the world. Defined in the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* as a depressive episode with peripartum onset, including onset during pregnancy and for up to 4 weeks postpartum [2], PPD is a clinical illness that often requires professional medical and/or psychological intervention. However, research shows that the point prevalence of depressive symptoms appears to peak between 2 and 7 months postpartum [3]. Due to this discrepancy, it is likely that many cases are missed.

In addition to the symptoms of depression, many women also experience comorbid symptoms, including those of generalized anxiety disorder, panic disorder, social phobia, and posttraumatic stress disorder [4]. In Canada, 10% of new mothers experience symptoms consistent with PPD, and an additional 8% experience PPD and anxiety symptoms together [5]. When combined, this figure is approximately equal to the worldwide prevalence rate of PPD symptoms, which has been calculated to be 17.7% [6]. As such, PPD affects a substantial number of Canadian women each year and, if left unaddressed, can have significant short- and long-term negative impacts on the mother, child, and family.

A recent systematic review assessing 122 studies found negative impacts of PPD on mothers' physical and mental health, infants' physical health and development, mother-infant bonding, and familial factors, including finances and relationships with partners [7]. Outside of the immediate family structure, PPD can continue to have severe consequences. An economic analysis of perinatal mental illnesses in the United Kingdom found that the economic burden of these illnesses was UK £6.6 billion (US $8.9 billion) for live births in 2013, based on productivity losses, health and social care use, and resultant impairments and needs of the child [8]. When broken down to costs per woman by condition, perinatal depression was determined to cost UK £75,728 (US $101,748) per woman with the condition in the same year, with UK £10,237 (US $13,753) of the cost paid by the public sector [8]. On the basis of our similar sociodemographic makeup and socialized health care system, it is likely that Canadian figures align greatly with this estimate. As such, ensuring that women have access to effective, acceptable, affordable, and usable treatment options is a matter of societal and economic importance, in addition to advancing the health and wellness of individuals and families.

Barriers to Accessing Care

There is a wealth of available management options for PPD, with a strong evidence base supporting the use of psychological approaches, including cognitive behavioral therapy (CBT) and interpersonal psychotherapy; psychosocial options, including peer support and nondirective counseling; and pharmacological treatments, specifically antidepressants [9]. Despite these options, a large proportion of women experiencing PPD do not seek or engage with these options, even when referred [10-12]. The literature has consistently shown that help-seeking behaviors in these women are impacted by a wide range of barriers, including social, instrumental, and structural barriers [13-15].

The main social barrier, which can have wide-ranging effects on all levels of help seeking, is stigma. In many cases, societal and cultural stigma lead to feelings of shame and guilt and feed into some common fears for women experiencing PPD, including child apprehension [13-15]. These feelings can then compromise disclosure. On a more personal level, a lack of social support from family and friends can also create a barrier to accessing care and contribute to instrumental barriers that exist, such as a lack of childcare to attend appointments [13-15].

Financial constraints are another major instrumental factor in help seeking [13-15]. A recent study in British Columbia (BC) found a significant association between socioeconomic status and access to health services, indicating a gap in care for women of low socioeconomic status who may already be underserved [16]. This is especially true for mental health services, as Canadian health care systems generally only cover psychiatric and community-based mental health services. The concentration of care options in urban centers can also be detrimental, as transportation and other indirect costs of accessing care are often identified as barriers to accessing care [13-15].

Finally, there are structural factors that impact access, with a lack of knowledge and information identified as a key element, both in patients and care providers [13-15]. This can lead to women remaining undiagnosed for longer than is necessary, and once these women are identified and directed to resources, they may be met with provider barriers, such as a lack of culturally safe care, and wait times that continue to impede their help seeking [13-15].

Throughout each sphere of life, women face barriers and gaps in accessing in-person care that contribute to and exacerbate their experiences of PPD. As such, there is a need to create and implement mental health management resources and interventions that address the inescapable stigma present in different cultural and social settings and that are effective, affordable, and accessible.

Uses of eHealth in Mental Health

The field of eHealth, involving the translation of psychological and other health interventions via telephone, apps, or web-based platforms, has been instrumental in the movement toward making health care accessible and personalized. By making information, education, and management resources readily available, eHealth allows individuals to "become partners in their own health" [17], empowering them to make decisions alongside their care providers.

eHealth interventions have been created and assessed for a wide range of common mental health conditions and are considered a cost-effective and accessible way to deliver care [18]. Several such interventions have been created using computerized CBT as the foundation of the provided therapeutic modules and have demonstrated efficacy in treating common mental health disorders, including depression and anxiety [18-21].

In addition, in recent years, there has been the introduction of multiple web-based interventions specifically for the treatment of perinatal mental health conditions. Two recent systematic reviews found that these interventions are moderately effective (effect size Hedges *g*=0.60) at improving maternal mood; however, both indicated a need for more research in this area to provide additional evidence and address existing limitations [22,23].

In the literature, accounts of these existing interventions for perinatal mental health are found around the world, covering a range of therapy types and modes of delivery. A group in the United Kingdom created Netmums, an intervention that provides guided behavioral activation treatment [24,25]. An intervention created jointly by groups in the United States and Australia provides a CBT-based program with personal coach assistance, called MomMoodBooster or MumMoodBooster, depending on the location [26-28]. A Swedish team created a CBT program based on an existing, established internet CBT program used in Sweden [29]. A Canadian group created a therapist-assisted CBT program [30]. In randomized controlled trials, each of these interventions demonstrated significant reductions in depressive symptoms following treatment compared with control groups [24,25,28-30].

However, only some of these interventions have published information about patient engagement during the design and development of these platforms, and there are still gaps in creating a truly user-centered tool. The Canadian therapist-assisted CBT program assessed qualitative feedback from clients who used the researcher-designed program during a randomized controlled trial and identified areas of improvement for the future [31]. As such, the resulting interface seems to be entirely designed and developed by the research team, with little involvement from end users. Meanwhile, the program content for Netmums was based on a previous qualitative inquiry using semistructured interviews, which assessed women’s needs and preferences for perinatal-specific CBT [32]. However, this work did not focus on an eHealth solution; therefore, there may still be unidentified gaps that come from employing technology for therapeutic purposes. Conversely, the MomMoodBooster program received significant qualitative feedback through the use of focus groups and usability testing to identify user needs specific to a web-based program [26]. However, these focus groups involved the presentation of preidentified content and user interface features, which were then adapted. As such, there may be missing features that could have been identified through a priori discussions.

As outlined in the established standards for evaluating eHealth interventions, a crucial time for evaluation is during the conceptualization and design phase [33]. However, research has shown that many eHealth interventions are only evaluated at the end of the study period, once the intervention has been fully developed and tested [34]. This study aims to inform the design and development of a proposed web-enabled intervention by first identifying the gaps in care and assessing the needs, values, and preferences of women experiencing PPD. As such, it employed the principles of integrated knowledge translation, which asserts that “involving knowledge users as equal partners alongside researchers will lead to research that is more relevant to, and more likely to be useful to, the knowledge users” [35]. Similarly, this aligns with the ideals of user-centered design, which is an iterative process of web design that involves end users throughout the development lifecycle [36]. A user-centered design establishes the need to know who end users are and to involve them early and often, to design tools that are relevant and useful [37]. A more significant inclusion of participants during the formative phases of content design and development would likely be beneficial in creating more applicable, effective interventions.

Objectives

This was the first stage of a multiphase study aiming to develop a web-enabled intervention that delivers psychosocial skills and education to women experiencing PPD. This preliminary phase had the following objectives: (1) to determine the unmet psychoeducational needs of non-Indigenous, nonmigrant women with PPD and (2) to explore how a web-enabled intervention could help to meet those needs, including what specific features and content this intervention must have for women to use it. Future phases of this study aim to collect similar data from both Indigenous and immigrant women.

Methods

Overview

As the first stage of a multiphase project, this study used focus groups to explore and understand the needs, preferences, and experiences of non-Indigenous, nonmigrant Canadian women who have experienced PPD to inform the development of a web-enabled intervention to support PPD management. This study will inform future studies assessing the needs of Indigenous women and immigrant women in BC to ensure that the intervention is accessible and acceptable to Canada’s diverse populations.

Study Population

To ensure that a diverse set of experiences and needs were included, a purposive sample of participants was recruited from 7 communities across BC (Table 1), varying in location, size, population demographics, and access to resources. These locations included Vancouver, Burnaby, Surrey, Kelowna, Prince George, Comox, and Victoria. In addition, women who lived outside of these specific communities, but who were interested in participating and lived in the general vicinity, were invited to attend focus groups in person or via Skype.

Table 1. Number of participants in each group location.

Location	Participants, n	Age (years), mean (SD)	Number of children, mean (SD)
Surrey	7	34.7 (5.6)	1.6 (0.5)
Vancouver	6	36.0 (3.5)	1.5 (0.5)
Burnaby	5	32.2 (4.4)	1.4 (0.5)
Prince George	2	34.0 (7.1)	2.0 (1.4)
Kelowna	5	31.0 (4.3)	2.2 (0.4)
Victoria	4	36.5 (4.9)	1.5 (0.6)
Comox	2	40.0 (5.7)	2.0 (1.4)
All locations	31	34.5 (4.9)	1.7 (0.6)

Eligibility

Eligibility criteria were set to ensure that focus group participants were able to provide insightful reflections on the helpfulness of mental health management strategies for PPD. To meet the inclusion criteria, participants were required to be 18 years or older; assigned female at birth and identify as a woman; identify as a non-Indigenous Canadian; be able to read, write, and speak conversational English; and have experienced PPD within the last 5 years but no longer meet the diagnostic criteria. All inclusion criteria were self-reported by participants. Before enrollment, participants were screened for current depressive symptomatology over the phone using the Edinburgh Postnatal Depression Scale (<12), which is the screening tool currently recommended by Perinatal Services BC [38].

Recruitment

Ethics approval was obtained from the University of British Columbia and BC Children's and Women's Hospital Research Ethics Board. Ethical considerations included providing all participants with a list of mental health resources available on the web and across the province of BC following the phone screen process and following each focus group because of the potential for triggering distress. In addition, a protocol was developed for participants who endorsed thoughts of self-harm, which ensured that they had access to mental health supports. Purposeful sampling was used to identify participants, with enrolled participants encouraged to aid in snowball recruitment through their own social networks. Study advertisements were promoted through social media and local service providers, with the majority of participants responding to ads on local Facebook groups, and through the Pacific Postpartum Support Society—a community partner in the Lower Mainland of BC, which encompasses the largest urban population in the province.

Procedures

Three distinct phases of participation were created, including completion of a demographic questionnaire, attendance at one focus group session, and an optional opportunity to take part in member checking through reviewing participants' own transcripts.

Demographic Questionnaire

A demographic questionnaire was completed by all enrolled participants; the questionnaire collected information relating to participants' personal life experiences, including data on socioeconomic status, intimate and family relationships, and medical history. Participants were offered the option to complete the questionnaire at home via an emailed unique survey link or mailed paper copy or at their focus group session on an iPad provided by the researchers. Study data were collected and managed using Research Electronic Data Capture tools hosted at BC Children's Hospital [39]. Descriptive statistics were analyzed using Statistical Package for the Social Sciences (version 26.0; IBM) [40].

Focus Group Interviews

As described earlier, focus groups were conducted in 7 communities across BC, with non-Indigenous, nonmigrant Canadian women. It is well established in the focus group methodology literature that sociocultural homogeneity within the group can be an important factor in creating a more comfortable and open sharing environment [41,42]. As such, Indigenous and immigrant focus groups were held separately in the second phase of this research.

Focus groups were held in local public venues, such as libraries and community centers. Skype participation was available for participants who were unable to attend a group in person because of unforeseen circumstances but who still wished to attend remotely. In these cases, Skype participants attended the same session as in-person participants but were displayed on a monitor for the in-person participants to see, with a camera setup so that the remote participant could also see the group. All study procedures for Skype participants were identical to those of in-person participants, via remote participation. Two participants chose to attend focus groups via Skype, and these participants were involved in the 2 rural focus group sessions in Prince George and Comox. These focus groups ranged from 1 to 2 hours in length, depending on the number of participants and time it took to thoroughly discuss emergent topics.

In addition to the study coordinator acting as a facilitator, each focus group employed an experienced cofacilitator from the local community to create a more accessible and acceptable environment for the participants. Cofacilitators were mainly service or care providers and included a public health nurse, a PhD student in clinical psychology, a family support worker for a pregnancy outreach program, the director of the Pacific Postpartum Support Society, and the study principal investigator who is a registered psychologist. These collaborators were recruited through general internet searches for local services for new parents and those who may be experiencing PPD. To ensure that any potential power dynamics between participants and providers were acknowledged, cofacilitators briefly discussed at the beginning of each focus group their role in the group and the community and encouraged participants to feel open to share their experiences, both good and bad, without fear of repercussion. In addition, included in the ground rules set out in the focus group interview guide to ensure participant safety and comfort, the facilitators would state the following: "There are no 'right' or 'wrong' answers, just different opinions. We ask that you respect each other's opinions and experiences, and say what is true for you without fear of judgement" (Multimedia Appendix 1).

The research team created a semistructured interview guide, which was not piloted before collecting study data. The interview guide posed a series of questions that transitioned from an open discussion of participants' experiences with PPD and available treatment options to more specific questions regarding their ideas for the design of the proposed web-enabled resource, including interactive features, peer support options, and the inclusion of partners. Each group concluded with a discussion of what participants regarded to be the most important factor in their personal recovery and what would be necessary for the research team to consider when designing the eventual platform. At least one focus group was run in each community, with no further groups run when saturation was reached, that is, no new ideas or information were observed [43].

In all groups except for one, because of time constraints, a discussion of an existing web-based CBT program called BounceBack was included, in addition to the regular interview guide, to inform the design and delivery of a potential CBT component in the proposed web-enabled intervention. BounceBack is a physician-recommended website available across Canada [44]. It has a significant evidence base to support its efficacy in managing a range of mild to moderate mental health disorders, including anxiety, depression, and eating disorders [45-47]. The group discussion of BounceBack was used to inform the design and delivery of CBT through the proposed web-enabled intervention and entailed an interactive walk-through of the website and a discussion of its strengths, weaknesses, and relevance to PPD.

To increase accessibility, women were encouraged to bring their young children if necessary, provided a catered meal during the group, and offered a $50 honorarium and parking reimbursement. Each session was audio recorded and transcribed verbatim by a contracted company, with additional context provided by field notes taken by a research assistant during each group. Before transcription, each employee at the transcription company was required to sign a confidentiality agreement, and participants were informed of this process to ensure that their informed consent was provided.

Member Checking

Participants were provided the optional opportunity to review their transcripts before analysis to honor participants' ownership of their experiences, how they were represented, and how they were interpreted to inform the resultant web-enabled intervention. If participants chose to take part in this step, they were sent the anonymized transcript from their focus group and informed of their study ID so that they could identify their individual comments. Participants were then asked to provide feedback on the accuracy of their transcript, any information that they would like redacted or clarified, and any unidentified portions of the transcript that were attributed to them, for example, if the transcriber was unable to determine who was speaking. Member checking substantiates the trustworthiness and transferability of results as participants confirm their dependability and credibility [48]. Of the 31 participants, 7 participated in member checking. Moreover, 4 of the 7 participants provided minor feedback related to clarifying words that were unintelligible or changing background information to be less identifiable. None of the changes made impacted the resultant analysis.

Qualitative Analysis

Anonymized, verbatim focus group transcripts were analyzed using the qualitative analysis software NVivo 12 (QSR International), which allowed for streamlined data management and synthesis across multiple research staff [49]. The 3 researchers who performed the analysis—all of whom were White, cis-gendered, nulliparous women—approached the data guided by a social constructivist framework, thus exploring and acknowledging how the data were shaped by who was in the room, opportunities to speak, and perceptions of safety in sharing during the focus groups [50].

The team used a qualitative descriptive approach to data analysis, identifying themes inductively and thematically. Qualitative description (QD) is often used in health research to inform the development of interventions or policies that can improve health outcomes for various populations [51]. On the basis of exploring "the who, what, and where of events and experiences" [52], QD provides a straight description based on participants' responses, making use of participants' own language to support the themes that emerge [52-54]. Its usefulness in informing interventions and its ability to assist in translating findings directly and rapidly to improving care makes QD a well-suited approach for the development of a web-enabled intervention [51]. Consistent with the literature for QD, inductive thematic analysis was used to explore and explain the focus group data [53]. Thematic analysis is described as a method of "identifying, analyzing, and reporting patterns" that exist within a qualitative data set [55].

In assessing the theoretical strength of qualitative research, there are a number of reputable frameworks, including the criteria developed by Lincoln and Guba [56] to assess trustworthiness, which include credibility, transferability, dependability, and confirmability. In this study, the development of the names, definitions, and relationships of codes, categories, and themes based on the approaches of Miles et al [57] were tracked via reflexive memos with illustrative quotes. In addition, each transcript was coded by a single investigator, and the resultant coding was then reviewed by a second investigator, with disagreements in coding discussed by the whole research team to arrive at the accepted analysis. This strengthened the dependability and trustworthiness of the results, as these practices facilitated the 3 investigators' intercoder reliability [58,59]; triangulation [60]; and arrival at a shared vision with participants, as confirmed via member checking [48]. When reporting participants' responses, their age, focus group location, and number of children were provided to contextualize the information and lent itself to the credibility, transferability, and dependability of the findings [61]. The creation and use of high-quality data is a significant step toward informing accessible, acceptable, and relevant care for women experiencing PPD across Canada.

Results

Demographics

A total of 31 women participated in the 7 focus group sessions (Table 2). The mean age was 34 years (range 24-44), with a mean of 2 children (range 1-3). This sample of women was predominantly White, highly educated, and married.

Table 2. Sociodemographic characteristics of focus group attendees (N=31).

Variable[a]	Value
Age (years), mean (SD)	34.5 (4.95)
Sexual orientation, n (%)	
Heterosexual	29 (94)
Bisexual	2 (6)
Ethnicity, n (%)	
Chinese	2 (6)
Hispanic or Latin American	1 (3)
South Asian (East Indian, Pakistani, Sri Lankan, etc)	1 (3)
White European	25 (81)
Biracial	2 (6)
Education, n (%)	
Graduated high school or earned GED[b]	2 (6)
Attended some college or university	4 (13)
Graduated 2-year college or university	8 (26)
Graduated 4-year college or university	9 (29)
Postgraduate degree	8 (26)
Employment, n (%)	
Full time	10 (32)
Part time	5 (16)
On maternity leave	5 (16)
Stay-at-home caregiver	7 (23)
Student	2 (6)
Unemployed	2 (6)
Annual household income, n (%)	
Less than Can $20,000 (US $15,634)	2 (6)
Between Can $20,000 and $39,999 (US $15,634 and $31,266)	2 (6)
Between Can $40,000 and $59,999 (US $31,267 and $46,900)	7 (23)
Between Can $60,000 and $79,999 (US $46,901 and $62,534)	3 (10)
Between Can $80,000 and $99,999 (US $62,353 and $78,167)	4 (13)
Between Can $100,000 and $119,999 (US $78,168 and $92,801)	2 (6)
Between Can $120,000 and $139,999 (US $92,802 and $109,434)	2 (6)
Between Can $140,000 and $159,999 (US $109,435 and $125,068)	2 (6)
More than Can $160,000 (US $125,069)	5 (16)
Prefer not to answer	2 (6)
Relationship status, n (%)	
Never married	1 (3)
Dating	1 (3)
In a relationship	1 (3)
Married	25 (81)
Common law	1 (3)
Separated	2 (6)

Variable[a]	Value
Number of children, mean (SD)	1.7 (0.65)
Number of PPD[c] diagnoses, mean (SD)	1.1 (0.30)

[a] Percentages may not equal 100 because of rounding.

[b] GED: general education development.

[c] PPD: postpartum depression.

Themes

Five major themes emerged: bridging gaps to meet needs; providing validation to combat stigma; nurturing capacity to cope, manage, and/or reach wellness; empowering people to take ownership over their mental health; and offering customization to ensure relevance. Each of these themes identifies a need (eg, combatting stigma) and a way to address that need (eg, providing validation).

Bridging Gaps to Meet Needs

Participants identified gaps that exist in knowledge about PPD and care options for PPD. They explained how this contributes to unmet needs, such as being unaware of the manifestations of PPD and finding care to be inaccessible. Opportunities for a web-enabled intervention to assist in bridging these gaps in the interests of meeting their needs were discussed.

The gaps described by participants ranged from access to care (ie, eligibility), care provision (ie, availability and acceptability), and awareness and understanding of postpartum mental health concerns. Multiple participants identified that the current definition of *postpartum* compromises responsiveness to their needs and contributes to their needs remaining unmet, as illustrated by the following participants:

> *...because one thing I have to say is the postpartum period is not a year. And that, that bothers me because I was in a race the first time for this to be done with. [...] I struggled for four years after my second.* [SUP21, 36, Vancouver, 2 children]
>
> *...so, I only saw (doctor) the one time. [...] Then when I called back and my son was already a year old, they said, "Well, it's too late." Like, I guess they only offer help for up to a year.* [SUP7, 39, Surrey, 2 children]

They explained how a web-enabled intervention could bridge these gaps:

> *I guess that seems like it might be a nice advantage of an app. Like, you wouldn't need to be one-year. Under one-year postpartum. You can just access it any time. And benefit from it.* [SUP28, 31, Burnaby, 1 child, 2 additional children as surrogate]

Even when eligible for care—for example, less than 1 year postpartum—many participants discussed the barriers they faced in accessing available options, such as the timing of appointments for care and the impact of long wait times to see specialists:

> *...I got referred to the group so often. It's like, "Yes, you need help. This is the help option," and I couldn't access that option. [...] I have friends that volunteered to call in sick so they could look after my son [tearful voice] so I could go, you know? And it's just that option wasn't there.* [SUP25, 42, Victoria, 2 children]
>
> *This was in November, and they weren't able to give me an appointment until March. So unfortunately, I wasn't diagnosed until March, like, officially, and that kind of, like, does break my heart and breaks my husband's heart.* [SUP12, 37, Burnaby, 1 child]

A web-enabled intervention was supported as a way to provide care to women who face these barriers, either as a first-line intervention, or as a supplement to in-person care:

> *[...] in my situation here, because resources are scarce [...] especially in northern, rural B.C., it's hard to access things, so whether that be a counsellor or sometimes even a good internet connection, right? So, I would need a resource to be as self-directed as possible.* [SUP13, 29, Prince George, 3 children]
>
> *[It] would be great for supplementing and helping a woman before they actually get to see a therapist or start on their medication or something and then would support them through that as well.* [SUP11, 32, Surrey, 2 children]

A pervasive topic across the focus groups was a lack of awareness and understanding regarding PPD. Awareness was defined as knowing of the existence of PPD as a mental health concern, whereas understanding went deeper into knowing *about* PPD, including symptoms and potential treatment options. Participants expressed how a lack of knowledge in their friends and family regarding the existence and presentation of PPD could contribute to unmet needs. A web-enabled intervention was endorsed as a way to engage and educate these support people to better assist women by providing information about PPD, ways to help, and how to access care:

> *Well, I think it's important for [partners] to be involved. I don't think that, they don't understand either. Like, they don't, I know mine wasn't. Like, he had no idea.* [SUP26, 29, Prince George, 1 child]
>
> *Or even signs, too, because I think, a lot of times, they notice that things are going on but don't necessarily know what to do with that information or what it's really reflecting, especially if they don't have any [experience], they've never even been exposed to information on postpartum depression.* [SUP32, 30, Kelowna, 2 children]

Although this investigation was focused on the role of a web-enabled intervention in mental health management, most participants expressed the value of its use as a preventive and promotive tool to feel better equipped in what to expect from the postpartum period:

> *I almost wish they made more of a case of it during, like, your pregnancy, you know? Like, "You should anticipate this. This could totally happen to you."* [SUP28, 31, Burnaby, 1 child, 2 additional children as surrogate]
>
> *I don't know if I would have been able to really look at anything while I was in it, but I think beforehand, like if it was available to people that are pregnant, then that might help [...] if I had already been on that app beforehand and, you know, knew what it was all about, then I would have been more open to that.* [SUP1, 32, Surrey, 1 child]

Unmet needs, such as a lack of knowledge regarding PPD, were thought to persist because of gaps that exist and are perpetuated by stigma, and participants have underscored how the normalization and validation of postpartum mental health is an effective way to combat stigma:

> *[...] normalization of the experience helps because most of us are sobbing in our homes, you know, thinking that this sucks. And we have nobody to tell, and there's guilt and shame and all this stuff that comes with it. And it's actually really common [laughter], um, but [...] it's supposed to be positive, right?* [SUP32, 30, Kelowna, 2 children]

Providing Validation to Combat Stigma

Participants spoke about their experiences with stigma, especially as it relates to notions of being a good mother. They described how it can become internalized and its effects on their mental health. For example, 1 woman expressed:

> *Like, for me, for example because I was feeling so, um, like, um, low, and like so hard on myself, feeling inadequate. You know, like, maybe I should be suffering, so when my husband said, "You should take care of yourself," it's like, "No. Like, I deserve to not feel good."* [SUP41, 31, Victoria, 1 child]

Validation of their feelings and experiences was key in addressing this internalized stigma, both through education and normalization of postpartum mental health concerns before they occur and through the sharing of stories from other women who have had PPD to fight feelings of isolation:

> *[...] in the support group, I never felt that. Like, everybody was saying their truth. It was like, "Oh, she is experiencing that. Wow. Okay. I am okay." There was no repercussion. It was just a safe place to, to feel okay.* [SUP8, 43, Surrey, 2 children]

Examples of how a web-enabled resource can integrate these features were provided, with 1 participant describing the inclusion of peer and professional encouragement:

> *Yeah, I think just some way of encouraging women whether it be through success stories or a connection with a counsellor, whatever it is, through the app that, um, would just encourage them not to give up and encourage them that it will get better.* [SUP30, 36, Comox, 1 child]

Another participant described what she found useful in the resources she used and how they could be adapted:

> *And seeing that stuff, I don't know if it was for you guys too, but, like, knowing that it's not, you're not crazy. It was just relief to know. I mean, when you read that, you're like, "Oh my gosh. This is totally what I'm going through." And it was so relieving, so having stuff on the app like that would be [good].* [SUP23, 39, Victoria, 1 child]

In addition to appreciating validation, participants described how they would want to reciprocate that validation by mobilizing their experiences and knowledge in the interest of demystifying and normalizing PPD, raising public awareness and combatting stigma:

> *Every chance I get to talk about my experience, I do. Like, I posted [...] on Facebook, and I asked everyone to share it. [...] I want someone to feel better, so I think just talking about it is important. Yeah, so you don't feel like a crazy person, right?* [SUP26, 29, Prince George, 1 child]

They identified that pervasive stigma toward PPD impacted attitudes expressed by family and friends, which was damaging to their mental health:

> *[...] so like, my mom would be like, "This is so weird. I have no idea what's wrong with you." Like, she'd actually say that to me. Yeah. She's like, "I loved my babies instantly. Like, as soon as I saw them, I was, like, in love." I'm like, "Oh God."* [SUP41, 31, Victoria, 1 child]

The effects of stigma were woven throughout personal and social interactions and greatly impacted the mental health and help seeking of women experiencing PPD. However, through these experiences, they discovered the power of validation and normalization to combat these attitudes. They also described lessons for nurturing capacity in other women to assist their journey to wellness:

> *I think it would be helpful in terms of normalizing it. Because I think the more resources via app or in person or whichever, helps to normalize it. Yeah, the more information you have, and the more, and right at your fingertips because you can't always leave the house. Right?* [SUP30, 36, Comox, 3 children]

Nurturing Capacity to Cope, Manage, and/or Reach Wellness

The need for tools to assist in management was clear when participants identified the challenges they faced in coping with and managing their mental health concerns and on their journey to wellness:

> *Yeah, I'd add, like CBT is great, but you have to have so many other things in place before it works. [...] I think a lot of people who don't have their basic practical needs met, like, "I don't actually have time*

> *to have a break because I have nobody to help me.”* [SUP10, 30, Vancouver, 1 child]
>
> *I knew that there was something really wrong, so I went, I think, once a week for support group and we did, my doctor tried to get me on meds, too, and I did not want to, so I just went to the support group. We learned meditation. We learned CBT, which was really helpful.* [SUP8, 43, Surrey, 2 children]
>
> *Also, I, like, had extreme panic attacks about taking the medication. I was so afraid [...] And so I worked with a therapist for probably four sessions before I even took one. [...] within a month, I was like, “Holy shit. I am like a complete, I love my child. I love my life. Wow, the sky is blue.” I was like I can’t believe I was so worried about taking medication that has completely changed my life and made me such a better mother.* [SUP15, 36, Vancouver, 1 child]

Many participants described how a web-enabled intervention could be used to nurture their capacity, thus providing the tools and education necessary to facilitate their ability to reach each of these levels of mental health management. Often, these tools were identified as being most helpful before mental health concerns arise, when they could be used as a preventive measure. However, once PPD occurs, many women emphasized the value of connecting them to peers with lived experience and providing evidence-based knowledge to deepen their understanding of PPD and the ability to determine their path to wellness:

> *I think there’s a lot of benefits to having both peers and professionals on it? Because you, you’re gonna trust a professional. They’ll tell you, “Drink 20 cups of water,” and you’re going to be downing the 18th one. As opposed to, like, a peer. A peer is a little bit more approachable, right? And you’re almost...you’re on the battlegrounds with them, so there’s a lot of validity in both.* [SUP19, 36, Vancouver, 1 child]
>
> *I was obsessed with the question, “Would I ever get better?” and I would ask everybody. [...] I would just plead with everybody, “Does this end?” So, to have success stories built into the app. Might be really helpful.* [SUP10, 30, Vancouver, 1 child]
>
> *[...] a community hub or a portal where, like, all that information is just. In one place. Because [...] I didn’t like reading about it. But the little bit I did read was from different sources, so if you just had everything in one place.* [SUP12, 37, Burnaby, 1 child]

They also discussed how gaining knowledge from peers and professionals can lead to personal insights and suggested ways of including opportunities for self-reflection to promote these insights within a web-enabled resource:

> *[...] mood tracking and stuff as just a way of, uh, paying attention to it, I think, was the point. Because a lot of the time when you’re depressed, you’re not...really thinking about the specifics of why you feel like you feel.* [SUP10, 30 Vancouver, 1 child]

Besides educating themselves, participants endorsed engaging family and friends to assist in nurturing their capacities to cope, manage, and reach wellness:

> *Yeah, I think anyone who might be involved, like someone, if they don’t have a partner or maybe if it’s their roommate or their mom or whoever it is that’s kind of closely tied to them. That could be that support partner.* [SUP30, 36, Prince George, 3 children]

Participants explained how having their experiences validated and their capacity nurtured empowered them to trust that they knew what their needs were and how best to meet them. However, they also discussed how having their capacity minimized and their needs dismissed by those they seek support from can leave them feeling disempowered and as though they have no control over their care:

> *[...] when I talked to the person [...] on the help line, um, [...] it was kind of strange, but I, because she almost sounded, made it sound like I was some kind of helpless victim of this thing, you know? But, which to me, didn’t sound like, [...] not very empowering [...] And I told her, “I am not a helpless victim. I know I can’t help the hormones and all of that, but there are things I can do.”* [SUP43, 44, Prince George, 3 children]

Empowering People to Take Ownership Over Their Mental Health

When feeling disempowered, participants often reflected on power dynamics within the health care system. Specifically, how their perceived lack of power was a detriment to their mental health and how a web-enabled intervention could promote agency in their pursuit of mental well-being:

> *I didn’t want to do medication, which is what I felt the psychiatrist just kind of pushed. She was very nice about it, but she kind of, I felt like she pushed it too hard.* [SUP2, 29, Surrey, 2 children]
>
> *[...] did I get help? I don’t think it was helpful. I think that the midwives thought it was helpful.* [SUP9, 39, Surrey, 2 children]
>
> *I find certain questions very intimidating because...what’s going to happen if I say, if I say I am having thoughts of hurting myself or someone else, it doesn’t mean I need to be slammed into a cage.* [SUP9, 39, Surrey, 2 children]

When describing how a web-enabled intervention could bridge these power dynamics, they discussed how it could supplement their relationships with care providers and ensure that they are able to make informed choices as partners in their own care:

> *[...] like, even if it was like, “Hey, if you’re in B.C., like, this is what you need to know in B.C.” Like, step by step, “Do this. Do this. Do this.”* [SUP12, 37, Burnaby, 1 child]

A common idea across focus groups centered on tracking your mental health using a web-enabled intervention to facilitate care:

> *Then yeah, if you are seeing someone [...] you can take out your app, and you can tell them, "Actually, I've noticed that over the last week, I've been more anxious. Let's talk about that" or "I've noticed that when I send my son to school, I start feeling like this."* [SUP11, 32, Surrey, 2 children]

In addition, they discussed the importance of their personal insights in determining the best pathway for them to achieve mental well-being. For example, identifying preferences in care options, such as self-care versus specialist care. Many identified self-care as a necessary way to make up for gaps in other care options:

> *[...] the piece that really turned the corner for me was [...] she drew a teapot, and she said [...] "You have all the things going into the tea kettle that are creating the pressure." So, she listed all of the things that I had talked about, and she said, "You have no outlet to let the pressure out, so and now your brain is exploding, basically." [...] And so when I feel that buildup of rage, because it usually expresses itself as rage, that's what I think about like, "My teapot is going to explode. How do I relieve the pressure?" So, I either take the pressure off myself and leave things that are putting pressure on or I go for a walk or I do something else that's going to help me to calm down. And I know what those things are, so I have my tools in my toolbelt, right?* [SUP25, 42, Victoria, 2 children]

Participants expressed how a web-enabled resource could support them in informing, understanding, and realizing their self-care practices:

> *I would like to see something that popped up, like a pop-up that said, "Self-care" [laughter]. Because this is a word I had no idea existed until postpartum.* [SUP8, 43, Surrey, 2 children]

Overall, participants considered how a web-enabled intervention could assist women by nurturing their capacity and empowering them to reach wellness on their own terms. They also identified how ensuring relevance is integral to these goals.

Offering Customization to Ensure Relevance

Participants suggested how a customizable interface through which people curate their journey toward mental well-being could offer relevance and make management more effective. Across themes, participants spoke about how current strategies and supports do not reflect their pregnancy and postpartum experiences. Consequently, the need to ensure relevance was highlighted throughout their discussions. Specifically, participants explained how a web-enabled intervention could offer customizable features to tailor resources to meet their needs:

> *I think that's a really important point for an app because it can't be one thing for all people. It needs to be, um, personalized to your, you know, what would help you the most.* [SUP24, 34, Victoria, 2 children]

On the basis of these customizations, women would then be able to access resources and support on their own terms:

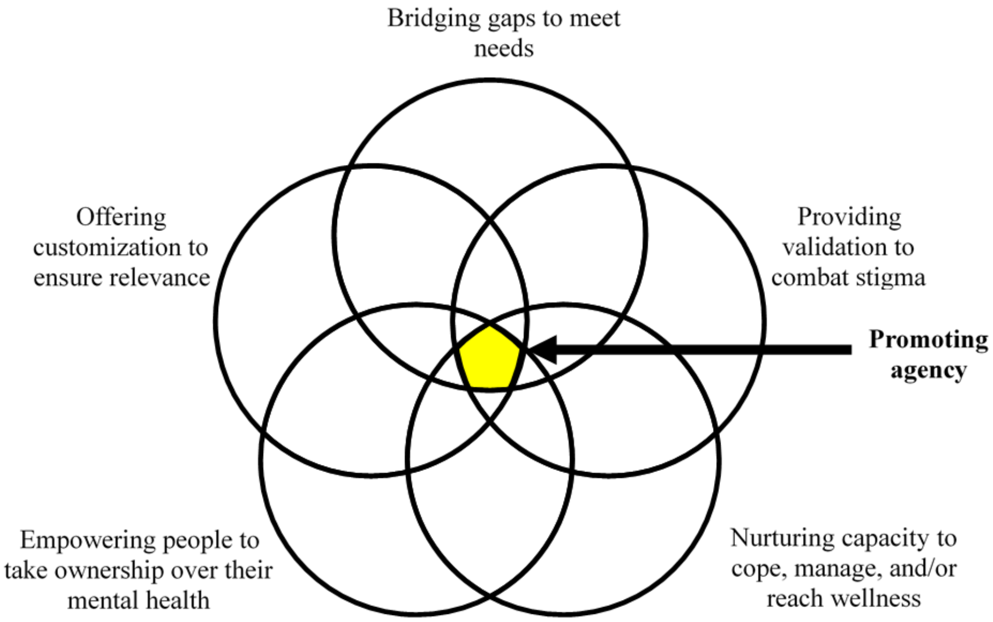
> *[...] just, like, a tiny bite, just a tiny bite because that's all I could have taken. And then maybe once I've taken that bite, there are, like, bigger bites going down. [...] so, like, little, tiny bread crumbs to be like, "Try this" or "Try that" and then, "If you try it and you like it, well here's a something a little bit different but the same. It might help you differently, better."* [SUP31, 32, Kelowna, 2 children]
>
> *Yes, I think maybe different levels, right? Different levels because sometimes you just need a mom that understands. And then sometimes you need to go to a suicide hotline.* [SUP1, 32, Surrey, 1 child]

They also reflected on their preferences for content delivery strategies and recommended interactive features. These contributions were often informed by their responsibilities and demands (eg, time constraints) as parents and what would be considered accessible and acceptable:

> *So even if it takes 20 pages, but 20 pages of smaller information because then it's easier to be like, "Okay. I was on page 5. I'll continue it later," rather than getting lost on a whole book [...] but something you can just quickly scroll through while you are breastfeeding, feeding a child, or anything, right?* [SUP2, 29, Surrey, 2 children]
>
> *You almost have to, like, you almost need to BuzzFeed it [laughter], you know? [...] I'd love to see, like, you know, "Three self-care tips you can do right now."* [SUP12, 37, Burnaby, 1 child]
>
> *I think, like, if you don't want to sit there and read for. Then you can watch the video. You can, like, pick and choose out what you want or what you think relates to you.* [SUP26, 29, Prince George, 1 child]

Ultimately, participants identified gaps and expressed how a web-enabled intervention could fill their needs for validation, increased capacity, and empowerment by offering relevant content. At the intersection of these 5 themes exists the overarching goal of promoting agency in women experiencing PPD so that they can determine and pursue the most effective and acceptable path to personal mental health and wellness (Figure 1).

Figure 1. Intersection of the 5 themes leading to the overarching goal of promoting agency.



Integration of Research Findings to Define Next Steps

To actualize the preferences of participants, it was necessary to explore specific recommendations for how their needs could be met through the use of a web-enabled intervention. Explicit examples of this were provided across the focus groups, with participants describing key features related to the emergent themes (Textbox 1). Moving forward into web development, the research team will integrate these features and content areas to ensure a user-centered, patient-oriented tool is created.

Textbox 1. Overview of how the emergent themes relate to specific features and content details.

Bridging gaps to meet needs

- Pathway to local services
- Self-directed therapies (cognitive behavioral therapy, mindfulness, etc)
- Availability throughout maternal trajectory

Providing validation to combat stigma

- Chat forum (with moderators)
- Peer buddy system
- Blog posts to share stories

Nurturing capacity to cope, manage, and/or reach wellness

- Resource hub
- Evidence-based information
- Resources for support people

Empowering people to take ownership over their mental health

- Daily prompts/check-ins
- Mental health tracking
- Self-care information
- Resources for holistic care

Offering customization to ensure relevance

- Profile/baseline assessment
- Options for opting in/out of features
- Content delivery for different learning styles (ie, written, videos, audio, etc)

Discussion

Making Connections

This study aims to determine the unmet needs that persist for women experiencing PPD and to explore how a web-enabled intervention for PPD could assist in meeting those needs. These findings clearly demonstrate that there is a range of unmet needs, including a need for education, validation, empowerment, and accessible care.

This reflects the current knowledge base regarding women's perceived barriers to accessing care and their unmet needs in PPD. As described throughout our themes, these gaps exist on multiple fronts and at multiple levels of mental health management. Stigma is consistently described as a significant barrier to help-seeking behavior [13-15]. As seen in this study, and as supported by previous literature [13], cultural stigma related to pregnancy and postpartum experiences can lead to guilt and shame in women. In addition, participants demonstrated how a lack of validation and normalization of postpartum mental health concerns exacerbates these feelings [62].

Woven throughout the themes is a lack of knowledge and understanding of postpartum mental health—another significant barrier to the pursuit of mental well-being [13-15]. This is directly related to stigma, validation, nurturing capacity, and empowerment, and much of what participants discussed in these themes involved feeling alone in their experiences, not having the information and tools to manage their mental health, and requiring education to become partners in their journey to mental well-being. A lack of knowledge regarding PPD in general and available resources and supports has been identified consistently in the literature, contributing to unmet needs [13-15].

Our themes clearly point to the importance of providing a resource that can be tailored to, and addresses, the diversity of needs and complexity of experiences faced by women experiencing PPD. In addition, ensuring that mental health resources are made accessible to women, regardless of factors such as location or income, is crucial.

Despite this sample of participants being predominantly well educated and affluent, the majority struggled to access care for PPD. However, financial constraints (which could be linked to income) and a lack of knowledge about PPD (which could be indirectly impacted by education level) are consistently highlighted as barriers to accessing care in the literature [13-15]. The financial aspect is especially important in Canada, where the universal health care system funds only psychiatric and community-based mental health services and pharmacological treatments [63]. This suggests that mental health concerns and care are prohibitive, even for those of affluent status. It also seems that social factors, such as stigma and social support, and structural factors, such as location, wait times, and health care provider knowledge, may play a strong role in facilitating access to care options for this population.

eHealth may provide critical access to information and services through its accessibility and affordability. In 2018, internet connectivity was available to 98% of British Columbians and reached target speeds of 50 Mbps+ in 93% of households in the province, meaning that it is widely available for personal use [64]. Although these numbers are reduced in rural and remote communities, many of these same communities have strong local public services, such as libraries or community centers, which provide free access to internet and technology. Although not discussed by our rural participants, future focus groups with Indigenous and immigrant women will continue to assess the accessibility of eHealth options to ensure that they are a feasible option for all potential users.

Finally, although out of the scope of this analysis, which pooled all 7 focus groups from across BC into one assessment, the geographic location expectedly played a role in the needs of women experiencing PPD. When discussing a web-enabled intervention, women based in urban areas—especially the Lower Mainland of BC—expressed a need to connect with local in-person resources, whereas women from rural and remote communities, who do not have in-person resources available, required a support that could be used as a stand-alone, first-line resource. Research has shown that women in rural settings have lower rates of PPD detection and treatment, which leads to health disparities in these populations [65], further bolstering the importance of exploring the diverse needs of women experiencing PPD to create an accessible and acceptable intervention that fills the gap.

Current eHealth Interventions for PPD

Patient engagement and acceptability are often overlooked in interventions using eHealth to treat perinatal mental health concerns. All existing interventions in this field rely solely on delivering psychotherapy via technology [24-30]. As evidenced by this study, this approach to management of PPD fails to address many of the key unmet needs of women. Some of these missed opportunities include providing education to address a lack of knowledge surrounding PPD, including peer support and evidence-based information, and incorporating partners in care to support women. These findings, along with recurring discussions of time constraints and lack of motivation that come with new parenthood and depression, indicate a need to shift away from providing a self-guided therapy-based intervention and to move toward a psychoeducational approach.

Psychoeducational eHealth interventions provide evidence-based information on a range of health topics and may be used for prevention, education, or treatment purposes [66]. A quick internet search will provide several psychoeducational eHealth resources for perinatal mental health; however, few of these resources have published information about their development or evaluation. The available literature highlights the low quality of current websites and apps for perinatal mental health, highlighting a serious need for evidence-based, patient-centered resources [67,68]. This has left an opportunity to design and develop an eHealth intervention that can truly meet the needs of women experiencing PPD, which this study will be used to inform.

Next Steps

Looking toward web development, the design of this resource will be guided by the needs and preferences expressed by participants, with end user engagement at each phase of research to ensure that the patient's voice is centered in all decision making. This will include a multidisciplinary advisory committee guiding web development, patient partners involved in discussions of privacy and security, and appropriately representative participant samples in each phase of future research. Our aim is to ensure that end users are able to provide feedback and guidance to co-design a resource that addresses their needs, although compromise may be necessary for the feasibility of some features, especially those with potential privacy and security threats. The next phase of this research will gather qualitative feedback from Indigenous and immigrant women to ensure that the developed web-enabled program is culturally safe and meets the needs of all Canadian women experiencing PPD. Following the integration of the new focus group data into the intervention content, usability testing will be undertaken before the intervention is tested in a randomized controlled trial to ensure its efficacy in reducing depressive symptomatology in women experiencing PPD. Future research should explore women's experiences of PPD in low- and middle-income countries to determine how this digital health tool could be adapted to provide accessible and affordable care in these settings.

Conclusions

This research provided a breadth and depth of insight into the needs of women experiencing PPD that has not been seen before in the literature, specifically as it relates to informing the development of an eHealth intervention. Participants described a multitude of unmet needs that persist because of varying gaps in care and knowledge; however, they also uncovered opportunities to bridge those gaps and meet their needs, in the interests of gaining ownership over their mental health journeys. These findings suggest that a web-enabled intervention is perceived as a welcome addition to supplement currently available resources for PPD and provide concrete recommendations on how it can best serve this purpose. Overall, this study provides an exciting foundation for creating an accessible, acceptable, and effective resource to assist in the treatment of PPD.

Limitations

Despite these findings and the many strengths of the study protocol, gaps remain in the knowledge base that this study was unable to fill. First, based on the difficulty of recruiting for focus groups, we were unable to assess the needs of extremely rural and remote communities, where care options are already scarce. In addition, as this phase of the research included solely non-Indigenous, nonmigrant Canadian women, the sample does not reflect the values and preferences of Canada's Indigenous and migrant populations. However, this study is part of a larger research project, which will assess the needs of Indigenous women and immigrant women who have experienced PPD, in the interests of representing and understanding the diverse needs of Canadian women.

In addition, the use of focus groups has some well-established limitations. As focus groups necessitate having multiple participants, ensuring recruitment and attendance can be difficult [69]. This study struggled with recruitment, resulting in some focus groups having only 2 attendees. In addition, a common phenomenon in focus group discussions is *group-think* or the conversation becoming biased because of dominant personalities and the need to fit in [69]. Further, some people tend to be overrepresented in the transcripts of focus group sessions, whereas others are silenced [69]. However, one of the main strategies to combat these limitations is to employ experienced, well-trained facilitators [69,70]. By including multiple facilitators in each focus group session—one who is extremely familiar with the literature and study and another who is familiar with the local community and available services for PPD—we were able to mitigate some of these potential outcomes. Despite these limitations, focus groups are thought to be extremely useful in projects that involve assessing the needs of a population or evaluating proposals because of the ease of quickly building off each other in discussions to form ideas. Moreover, for this particular population of women, who are often isolated and ashamed of their experiences, the opportunity to meet and speak with peers to share their stories was expressed as being very therapeutic.

Acknowledgments

This research would not have been possible without a generous donation from The Gillespie Family Foundation, coordinated by the BC Women's Health Foundation. Additional funding was provided through the incredible fundraising of the SHOPPERS. LOVE.YOU. Run for Women, organized by Shoppers Drug Mart and the BC Women's Health Foundation. ML's graduate student salary was also supported in part by the Canadian Graduate Scholarship—Master's Award through the Canadian Institutes of Health Research. The authors would like to acknowledge the intellectual contributions of Dr Dorothy Shaw and feedback from Dr Simone Vigod. The authors would also like to acknowledge Kina Ellison and Raysa Camara for their assistance with transcription and data quality assurance.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group interview guide.
[DOCX File , 21 KB - jmir_v23i1e18934_app1.docx]

References

1. Oates M. Perinatal psychiatric disorders: a leading cause of maternal morbidity and mortality. Br Med Bull 2003;67:219-229. [doi: 10.1093/bmb/ldg011] [Medline: 14711766]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. Arlington, VA: American Psychiatric Publishing; 2013.
3. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005 Nov;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
4. Wisner KL, Sit DK, McShea MC, Rizzo DM, Zoretich RA, Hughes CL, et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. JAMA Psychiatry 2013 May;70(5):490-498 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.87] [Medline: 23487258]
5. Maternal Mental Health in Canada (Statistics Canada Infographics Catalogue No. 11-627-m). Statistics Canada. 2019. URL: https://www150.statcan.gc.ca/n1/en/pub/11-627-m/11-627-m2019041-eng.pdf?st=3CbcYwfP [accessed 2020-05-08]
6. Hahn-Holbrook J, Cornwell-Hinrichs T, Anaya I. Economic and health predictors of national postpartum depression prevalence: a systematic review, meta-analysis, and meta-regression of 291 studies from 56 countries. Front Psychiatry 2017;8:248 [FREE Full text] [doi: 10.3389/fpsyt.2017.00248] [Medline: 29449816]
7. Slomian J, Honvo G, Emonts P, Reginster J, Bruyère O. Consequences of maternal postpartum depression: a systematic review of maternal and infant outcomes. Womens Health (Lond) 2019;15:1745506519844044 [FREE Full text] [doi: 10.1177/1745506519844044] [Medline: 31035856]
8. Bauer A, Knapp M, Parsonage M. Lifetime costs of perinatal anxiety and depression. J Affect Disord 2016 Mar 1;192:83-90. [doi: 10.1016/j.jad.2015.12.005] [Medline: 26707352]
9. Stewart DE, Vigod SN. Postpartum depression: pathophysiology, treatment, and emerging therapeutics. Annu Rev Med 2019 Jan 27;70:183-196. [doi: 10.1146/annurev-med-041217-011106] [Medline: 30691372]
10. Prevatt B, Desmarais SL. Facilitators and barriers to disclosure of postpartum mood disorder symptoms to a healthcare provider. Matern Child Health J 2018 Jan;22(1):120-129. [doi: 10.1007/s10995-017-2361-5] [Medline: 28766092]
11. Fonseca A, Gorayeb R, Canavarro MC. Women s help-seeking behaviours for depressive symptoms during the perinatal period: socio-demographic and clinical correlates and perceived barriers to seeking professional help. Midwifery 2015 Dec;31(12):1177-1185. [doi: 10.1016/j.midw.2015.09.002] [Medline: 26433622]
12. Kim JJ, La Porte LM, Corcoran M, Magasi S, Batza J, Silver RK. Barriers to mental health treatment among obstetric patients at risk for depression. Am J Obstet Gynecol 2010 Mar;202(3):312.e1-312.e5. [doi: 10.1016/j.ajog.2010.01.004] [Medline: 20207252]
13. Jones A. Help seeking in the perinatal period: a review of barriers and facilitators. Soc Work Public Health 2019;34(7):596-605. [doi: 10.1080/19371918.2019.1635947] [Medline: 31242074]
14. Button S, Thornton A, Lee S, Shakespeare J, Ayers S. Seeking help for perinatal psychological distress: a meta-synthesis of women's experiences. Br J Gen Pract 2017 Oct;67(663):e692-e699 [FREE Full text] [doi: 10.3399/bjgp17X692549] [Medline: 28847773]
15. Dennis C, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536X.2006.00130.x] [Medline: 17150072]
16. Hanley GE, Park M, Oberlander TF. Socioeconomic status and treatment of depression during pregnancy: a retrospective population-based cohort study in British Columbia, Canada. Arch Womens Ment Health 2018 Dec;21(6):765-775. [doi: 10.1007/s00737-018-0866-6] [Medline: 29860622]
17. Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. Int J Med Inform 2001 Apr;61(1):1-10. [doi: 10.1016/s1386-5056(00)00130-1] [Medline: 11248599]
18. Barr Taylor C, Luce K. Computer- and internet-based psychotherapy interventions. Curr Dir Psychol Sci 2003 Feb 1;12(1):18-22 [FREE Full text] [doi: 10.1111/1467-8721.01214]
19. Grist R, Cavanagh K. Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. J Contemp Psychother 2013 Sep 4;43(4):243-251 [FREE Full text] [doi: 10.1007/s10879-013-9243-y]
20. Foroushani PS, Schneider J, Assareh N. Meta-review of the effectiveness of computerised CBT in treating depression. BMC Psychiatry 2011 Aug 12;11:131. [doi: 10.1186/1471-244X-11-131] [Medline: 21838902]
21. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
22. Lee EW, Denison FC, Hor K, Reynolds RM. Web-based interventions for prevention and treatment of perinatal mood disorders: a systematic review. BMC Pregnancy Childbirth 2016 Feb 29;16:38 [FREE Full text] [doi: 10.1186/s12884-016-0831-1] [Medline: 26928898]
23. Loughnan SA, Joubert AE, Grierson A, Andrews G, Newby JM. Internet-delivered psychological interventions for clinical anxiety and depression in perinatal women: a systematic review and meta-analysis. Arch Womens Ment Health 2019 Dec;22(6):737-750. [doi: 10.1007/s00737-019-00961-9] [Medline: 31101993]

24. O'Mahen HA, Woodford J, McGinley J, Warren FC, Richards DA, Lynch TR, et al. Internet-based behavioral activation--treatment for postnatal depression (Netmums): a randomized controlled trial. J Affect Disord 2013 Sep 25;150(3):814-822. [doi: 10.1016/j.jad.2013.03.005] [Medline: 23602514]
25. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]
26. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. Web-based intervention for postpartum depression: formative research and design of the MomMoodBooster program. JMIR Res Protoc 2012 Nov 22;1(2):e18 [FREE Full text] [doi: 10.2196/resprot.2329] [Medline: 23612274]
27. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. MomMoodBooster web-based intervention for postpartum depression: feasibility trial results. J Med Internet Res 2013 Nov 4;15(11):e242 [FREE Full text] [doi: 10.2196/jmir.2876] [Medline: 24191345]
28. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 7;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
29. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
30. Pugh NE, Hadjistavropoulos HD, Dirkse D. A randomised controlled trial of therapist-assisted, internet-delivered cognitive behavior therapy for women with maternal depression. PLoS One 2016;11(3):e0149186 [FREE Full text] [doi: 10.1371/journal.pone.0149186] [Medline: 26930488]
31. Pugh NE, Hadjistavropoulos HD, Hampton AJD, Bowen A, Williams J. Client experiences of guided internet cognitive behavior therapy for postpartum depression: a qualitative study. Arch Womens Ment Health 2015 Apr;18(2):209-219. [doi: 10.1007/s00737-014-0449-0] [Medline: 25109484]
32. O'Mahen H, Fedock G, Henshaw E, Himle J, Forman J, Flynn H. Modifying CBT for perinatal depression: What do women want?: A qualitative study. Cogn Behav Pract 2012 May;19(2):359-371 [FREE Full text] [doi: 10.1016/j.cbpra.2011.05.005]
33. Eng TR, Gustafson DH, Henderson J, Jimison H, Patrick K. Introduction to evaluation of interactive health communication applications. Science panel on interactive communication and health. Am J Prev Med 1999 Jan;16(1):10-15. [doi: 10.1016/s0749-3797(98)00107-x] [Medline: 9894549]
34. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of ehealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
35. Guide to Knowledge Translation Planning at CIHR: Integrated and End-of-Grant Approaches. Ottawa, ON: Canadian Institutes of Health Research; 2015.
36. User-centered Design Basics. US Department of Health and Human Services. URL: https://www.usability.gov/what-and-why/user-centered-design.html [accessed 2020-06-28]
37. Still B, Crane K. UCD principles. In: Fundamentals of User-Centered Design: A Practical Approach. Boca Raton, FL: Taylor & Francis Group; 2017:43-65.
38. Edinburgh Postnatal Depression Scale (EPDS). Perinatal Services BC. 2019. URL: http://www.perinatalservicesbc.ca/health-professionals/professional-resources/health-promo/edinburgh-postnatal-depression-scale-(epds) [accessed 2019-09-01]
39. REDCap. URL: https://projectredcap.org/software/ [accessed 2018-09-01]
40. IBM. URL: https://www.ibm.com/downloads/cas/YDWQGBDL [accessed 2019-06-01]
41. Morgan D. Planning and research design for focus groups. In: Focus Groups as Qualitative Research. Second Edition. Thousand Oaks, CA: SAGE Publications; 1997:34-35.
42. Ivanoff SD, Hultberg J. Understanding the multiple realities of everyday life: basic assumptions in focus-group methodology. Scand J Occup Ther 2006 Jun;13(2):125-132. [doi: 10.1080/11038120600691082] [Medline: 16856469]
43. Saumure K, Given L. Data saturation. In: Given L, editor. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, CA: SAGE Publications; 2008:196.
44. BounceBack. Canadian Mental Health Association. 2019. URL: https://bouncebackbc.ca/ [accessed 2019-09-01]
45. Pittaway S, Cupitt C, Palmer D, Arowobusoye N, Milne R, Holttum S, et al. Comparative, clinical feasibility study of three tools for delivery of cognitive behavioural therapy for mild to moderate depression and anxiety provided on a self-help basis. Ment Health Fam Med 2009 Sep;6(3):145-154 [FREE Full text] [Medline: 22477905]
46. Williams C, Wilson P, Morrison J, McMahon A, Walker A, Andrew W, et al. Guided self-help cognitive behavioural therapy for depression in primary care: a randomised controlled trial. PLoS One 2013;8(1):e52735 [FREE Full text] [doi: 10.1371/journal.pone.0052735] [Medline: 23326352]
47. Lau M, Davis S. Evaluation of a cognitive behavior therapy program for BC primary care patients with mild to moderate depression with or without anxiety: Bounce Back, 2008-2014. BC Med J 2019;61(1):25-32 [FREE Full text]
48. Sandelowski M. Member check. In: Given L, editor. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, CA: SAGE Publications; 2008:502.

49. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/about/nvivo [accessed 2019-03-01]
50. Creswell J, Poth C. Qualitative Inquiry & Research Design, Choosing Among Five Approaches. Fourth Edition. Thousand Oaks, CA: SAGE Publications; 2017.
51. Sullivan-Bolyai S, Bova C, Harper D. Developing and refining interventions in persons with health disparities: the use of qualitative description. Nurs Outlook 2005;53(3):127-133. [doi: 10.1016/j.outlook.2005.03.005] [Medline: 15988449]
52. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
53. Kim H, Sefcik JS, Bradway C. Characteristics of qualitative descriptive studies: a systematic review. Res Nurs Health 2017 Feb;40(1):23-42 [FREE Full text] [doi: 10.1002/nur.21768] [Medline: 27686751]
54. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
55. Braun V, Clarke V. Using thematic analysis in psychology. Qual Health Res 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
56. Lincoln Y, Guba E. Establishing trustworthiness. In: Naturalistic Inquiry. Newbury Park, CA: SAGE Publications; 1985:289-332.
57. Miles M, Huberman A, Saldaña J. Fundamentals of qualitative data analysis. In: Qualitative Data Analysis A Methods Sourcebook. 3rd Ed. Thousand Oaks, CA: SAGE Publications; 2014:69-105.
58. Lombard M, Bracken C, Snyder-Duch J. Intercoder reliability. In: Allen M, editor. The SAGE Encyclopedia of Communication Research Methods. Thousand Oaks, CA: SAGE Publications; 2017:722-724.
59. van den Hoonaard WC. Inter- and intracoder reliability. In: Given L, editor. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, CA: SAGE Publications; 2008:446.
60. Rothbauer P. Triangulation. In: Given L, editor. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, CA: SAGE Publications; 2008:893-894.
61. Groenewald T. Memos and memoing. In: Given L, editor. The Sage Encyclopedia of Qualitative Research Methods. Thousand Oaks, CA: SAGE Publications; 2008:506.
62. Thomas LJ, Scharp KM, Paxman CG. Stories of postpartum depression: exploring health constructs and help-seeking in mothers' talk. Women Health 2014;54(4):373-387. [doi: 10.1080/03630242.2014.896442] [Medline: 24617862]
63. Sunderland A, Findlay L. Perceived need for mental health care in Canada: results from the 2012 Canadian community health survey-mental health (Statistics Canada, Catalogue No. 82-003-X). Health Rep 2013;24(9):3-9 [FREE Full text]
64. Communications Monitoring Report 2019. Ottawa, ON: Canadian Radio-television and Telecommunications Commission; 2019:278.
65. Geier ML, Hills N, Gonzales M, Tum K, Finley PR. Detection and treatment rates for perinatal depression in a state Medicaid population. CNS Spectr 2015 Feb;20(1):11-19. [doi: 10.1017/S1092852914000510] [Medline: 25307282]
66. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009 Dec 16;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
67. Zhang MW, Ho RC, Loh A, Wing T, Wynne O, Chan SW, et al. Current status of postnatal depression smartphone applications available on application stores: an information quality analysis. BMJ Open 2017 Nov 14;7(11):e015655 [FREE Full text] [doi: 10.1136/bmjopen-2016-015655] [Medline: 29138195]
68. Kirby PL, Reynolds KA, Walker JR, Furer P, Pryor TA. Evaluating the quality of perinatal anxiety information available online. Arch Womens Ment Health 2018 Dec;21(6):813-820. [doi: 10.1007/s00737-018-0875-5] [Medline: 29931445]
69. Wong LP. Focus group discussion: a tool for health and medical research. Singapore Med J 2008 Mar;49(3):256-60; quiz 261 [FREE Full text] [Medline: 18363011]
70. Leung F, Savithiri R. Spotlight on focus groups. Can Fam Physician 2009 Feb;55(2):218-219 [FREE Full text] [Medline: 19221085]

Abbreviations

BC: British Columbia
CBT: cognitive behavioral therapy
PPD: postpartum depression
QD: qualitative description

Edited by T Rashid Soron; submitted 06.04.20; peer-reviewed by M Park, L McCann, H O'Mahen, A Nguyen; comments to author 03.06.20; revised version received 14.07.20; accepted 19.08.20; published 06.01.21.

Please cite as:
Lackie ME, Parrilla JS, Lavery BM, Kennedy AL, Ryan D, Shulman B, Brotto LA
Digital Health Needs of Women With Postpartum Depression: Focus Group Study
J Med Internet Res 2021;23(1):e18934
URL: https://www.jmir.org/2021/1/e18934
doi:10.2196/18934
PMID:33404506

Original Paper

Uptake of Proactively Offered Online and Telephone Support Services Targeting Multiple Health Risk Behaviors Among Vocational Education Students: Process Evaluation of a Cluster Randomized Controlled Trial

Prince Atorkey[1,2,3,4], BSc, MPhil; Christine Paul[1,3,4], PhD; Billie Bonevski[1,3], PhD; John Wiggers[1,2,3,4], PhD; Aimee Mitchell[1,2], BPsych (Hons); Emma Byrnes[1,3], BSocSc (Hons); Christophe Lecathelinais[1,2,3], DESS; Flora Tzelepis[1,2,3,4], PhD

[1]School of Medicine and Public Health, Faculty of Health and Medicine, University of Newcastle, Newcastle, Australia
[2]Hunter New England Population Health, Hunter New England Local Health District, Newcastle, Australia
[3]Hunter Medical Research Institute, University of Newcastle, Newcastle, Australia
[4]Priority Research Centre for Health Behaviour, Faculty of Health and Medicine, University of Newcastle, Newcastle, Australia

Corresponding Author:
Prince Atorkey, BSc, MPhil
School of Medicine and Public Health
Faculty of Health and Medicine
University of Newcastle
University Drive
Callaghan
Newcastle, 2308
Australia
Phone: 61 0249246565
Fax: 61 0249246490
Email: prince.atorkey@uon.edu.au

Abstract

Background: A high proportion of vocational education students smoke tobacco, have inadequate nutrition (ie, low fruit and vegetable intake), drink alcohol at risky levels, or are physically inactive. The extent to which vocational education students will sign up for proactively offered online and telephone support services for multiple health risk behaviors is unknown.

Objective: The aim of this study is to examine the uptake of proactively offered online and telephone support services for smoking, nutrition, alcohol consumption, and physical activity risk behaviors, individually and in combination, among vocational education students in the Technical and Further Education (TAFE) setting. The characteristics associated with the uptake of online or telephone services for smoking, nutrition, alcohol consumption, and physical activity risk behaviors were also examined.

Methods: Vocational education students enrolled in a TAFE class in New South Wales, Australia, which ran for 6 months or more, were recruited to participate in a cluster randomized controlled trial from May 2018 to May 2019. In the intervention arm, participants who did not meet the Australian health guidelines for each of the smoking, nutrition, alcohol consumption, and physical activity risk behaviors were provided electronic feedback and proactively offered online and telephone support services. Uptake of support was measured by whether participants signed up for the online and telephone services they were offered.

Results: Vocational education students (N=551; mean age 25.7 years, SD 11.1; 310/551, 56.3% male) were recruited into the intervention arm. Uptake of the proactive offer of either online or telephone services was 14.5% (59/406) for fruit and vegetables, 12.7% (29/228) for physical activity, 6.8% (13/191) for smoking, and 5.5% (18/327) for alcohol use. Uptake of any online or telephone service for at least two health behaviors was 5.8% (22/377). Participants who were employed (odds ratio [OR] 0.10, 95% CI 0.01-0.72) and reported not being anxious (OR 0.11, 95% CI 0.02-0.71) had smaller odds of signing up for online or telephone services for smoking, whereas participants who reported not being depressed had greater odds (OR 10.25, 95% CI 1.30-80.67). Participants who intended to change their physical activity in the next 30 days had greater odds (OR 4.01, 95% CI 1.33-12.07) of signing up for online or telephone services for physical activity. Employed participants had smaller odds (OR 0.18, 95% CI 0.06-0.56) of signing up for support services for at least two behaviors.

Conclusions: Although the uptake of proactively offered online and telephone support services is low, these rates appear to be higher than the self-initiated use of some of these services in the general population. Scaling up the proactive offer of online and telephone services may produce beneficial health outcomes.

Trial Registration: Australian New Zealand Clinical Trials Registry: ACTRN12618000723280; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375001.

(J Med Internet Res 2021;23(1):e19737) doi:10.2196/19737

KEYWORDS

uptake; proactive offer; online support services; telephone support services; multiple health risk behaviors; vocational education students

Introduction

Vocational Education Settings and Multiple Health Risk Behaviors

Vocational education settings are mandated to offer education programs that equip students with technical qualifications that prepare them for specific occupations such as carpentry, hairdressing, plumbing, floristry, automotive engineering, and welding [1]. A high proportion of vocational education students compared with university students report tobacco smoking, poor nutrition, risky alcohol consumption, and physical inactivity [2,3]. For instance, in an Australian study, 35% of vocational education students reported smoking tobacco daily, 96% did not consume at least five serves of vegetables daily, 50% did not meet the recommended daily fruit consumption of at least two servings, 49% consumed alcohol at risky levels, 88% were physically inactive, and 98% engaged in multiple health risk behaviors [2]. It is important to target smoking, nutrition, alcohol consumption, and physical activity health risk behaviors because they are associated with an increased risk of chronic diseases and death [4]. The risk of chronic disease and mortality is higher when one engages in multiple smoking, nutrition, alcohol consumption, and physical activity risk behaviors [4,5]. Vocational education settings are predominately made up of young adults [2]; therefore, offering interventions that target these risk behaviors at an early age could reduce chronic diseases, mortality, and health care costs.

Smoking, nutrition, alcohol consumption, and physical activity health risk behaviors have been found to co-occur or cluster together in vocational education students [6-8]. For example, a study with community college students identified 3 clusters: *active, binge-drinkers with poor dietary intake*; *nonactive, moderate-smokers and nondrinkers with poor dietary intake*; and *moderately active, nonsmokers and nondrinkers with poor dietary intake* [7]. This highlights the importance of taking a holistic approach to behavioral interventions and targeting multiple health risk behaviors collectively. Transfer theory [9] also suggests that the advantages of addressing multiple health risk behaviors include that the knowledge and experiences acquired to successfully modify one behavior can be applied to change other health behaviors [10]. Given the rates of smoking, nutrition, alcohol consumption, and physical activity risk behaviors among vocational education students, and that such behaviors tend to cluster together, connecting this population to services that provide effective support to modify these health behaviors has the potential to improve health outcomes in this population.

Effectiveness of Online and Telephone Support Services in Modifying Health Risk Behaviors

Online and telephone interventions are effective in reducing smoking [11,12], improving nutrition [13,14], reducing alcohol consumption [15], increasing physical activity [13,14], and modifying multiple health risk behaviors [13,16], and they provide a means for changing these behaviors among vocational education students. Advantages of telephone and online interventions include privacy and reduction in stigma [17-19], avoidance of travel time and cost [20], potential to reach a large number of people [21,22], ability to reach individuals who may not seek support [23], and capacity to provide support when it is needed most [19].

In the United States and Australia, for example, evidence-based online and telephone interventions designed to modify health risk behaviors are available to the general population at no cost to users [24-26]. Online interventions available for the general population in Australia include QuitCoach (smoking) [27], Tertiary Health Research Intervention Via Email (THRIVE; alcohol) [28], Healthy Eating Quiz (fruit and vegetable consumption), and 10,000 steps (physical activity) programs [29]. Telephone services available to the Australian general population include Quitline (smoking) [30], Alcohol Drug Information Service (ADIS; alcohol), and Get Healthy Information and Coaching Service (GHICS; fruit and vegetable consumption and physical activity) [31]. Despite the efficacy of online and telephone support services in improving health risk behaviors, the uptake of such services is low [32,33]. For example, only 3% to 4% of smokers in Australia and the United States use the Quitline [32]. Less than 1% of adults in New South Wales, Australia, who are overweight and obese call the GHICS to receive support to increase their fruit and vegetable consumption and physical activity [34]. Only 0.3% of smokers reported using an internet-based intervention in an attempt to quit smoking in the past year [35]. Furthermore, when the figures provided in an Australian study were divided by the total Australian population aged 16 years and older [36] with access to the online service, only 0.4% were found to use the Healthy Eating Quiz [37].

The low rates of use of online and telephone support services may partially be due to the passive recruitment approaches (eg, mass media campaigns) used by these services, which generally rely on individuals to initiate contact with the service providers themselves [38]. Alternatively, proactive recruitment involves service providers initiating contact with the individual to offer support services and has been found to increase the uptake of such services [22,39-42]. For instance, 52% of smokers who were proactively offered Quitline telephone support accepted it [30], and 23% of people who were proactively offered the GHICS accepted it [43]. McClure et al [44] reported that 7% of smokers who were proactively invited to use an internet-based intervention (ie, Project Quit) accepted and visited the website and 4% signed up to use the service.

Factors Associated With Uptake of Support Services

Sociodemographic characteristics are associated with whether people will sign up to use online or telephone interventions. Skov-Ettrup et al [22] found that participants who spent 11 years or more in school compared with those who spent shorter periods in school were less likely to use telephone counseling to quit smoking. Furthermore, participants aged 18 to 39 years were more likely to sign up for an internet-based intervention targeting smoking than those aged 40 years or more [22]. A study conducted in Hong Kong also found that being employed and belonging to a middle-income family was significantly associated with accepting to use a telephone service to quit smoking [45]. Furthermore, Schneider et al [41] found that Dutch adults who were younger, male, highly educated, employed, and were in a relationship were more likely to express interest in the proactively offered online intervention targeting multiple lifestyle behaviors.

Existing studies on the uptake of online and telephone services and factors associated with uptake focus only on the general population or university students [22,45]. Only one study examined the uptake of online support services for multiple health risk behaviors [41], whereas the other studies focused on a single health risk behavior only (eg, smoking) [22,32,44,45]. Despite their high rates of smoking, nutrition, alcohol consumption, and physical activity risk behaviors, it is unknown whether vocational education students will sign up for proactively offered online and/or telephone support to modify one or more of these behaviors. Furthermore, whether the sociodemographic and psychological characteristics of vocational education students are associated with signing up for online and telephone support services is unexplored. Although no previous research has examined whether psychological characteristics are associated with uptake of online and telephone support services targeting multiple health risk behaviors, mental health issues have been found to co-occur or cluster with health risk behaviors [6,46]. Therefore, it is important to examine whether those with symptoms of psychological distress (ie, anxiety and depression) will sign up for support services targeting their health risk behaviors.

Objectives

The aims of this study are to examine the following among vocational education students in the Technical and Further Education (TAFE) setting:

1. Uptake of proactively offered online and telephone support services for each smoking, nutrition, alcohol consumption, and physical activity risk behavior and multiple health behaviors (ie, at least two behaviors).
2. Sociodemographic and psychological characteristics associated with the uptake of online and telephone support services for each smoking, nutrition, alcohol consumption, and physical activity risk behavior and multiple health behaviors (ie, at least two behaviors).

Methods

Study Design and Setting

The data used in this study were gathered from May 2018 to May 2019 as part of a cluster randomized controlled trial examining the effectiveness of electronic feedback and online and telephone support services targeting multiple health risk behaviors among vocational education students. Of the 14 TAFE campuses, 8 campuses were randomized to the intervention group and 6 campuses to the control group. The control campuses received no intervention. The data from all 8 intervention campuses, which were located in the Hunter, Upper Hunter, and Central Coast regions of New South Wales, Australia, are included in this paper. Vocational education students who attended these 8 intervention campuses were proactively offered online and telephone support services for those behaviors where they did not meet Australian health guidelines. The University of Newcastle Human Research Ethics Committee granted ethics approval for this study. Written approval was also provided by the Chief Education and Training Officer from TAFE New South Wales to conduct this research. The trial was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12618000723280).

Participants

Participants were eligible to participate if they were aged 16 years and above and enrolled in a TAFE class that ran for 6 months or more. Participants were excluded if they met the recommended health guidelines for all smoking, nutrition, alcohol consumption, and physical activity risk behaviors and were not able to read or write English.

Procedure

The TAFE service coordinator on each campus identified all eligible classes that ran for 6 months or more and approached the head teachers of the department followed by the teachers of the classes to obtain approval to attend those classes. In those classes where the head teacher and teacher granted approval, TAFE students were given an information letter in class about the research 1 week before data collection. On the day of data collection, members of the research team attended the class to administer the online survey. First, a verbal explanation of the study was provided to the class by a researcher, and interested students were given a computer tablet to complete the online questionnaire. Informed consent was obtained online on the first page before the online survey questions. Interested students selected that they would like to participate in the trial before the survey questions appeared on screen.

Intervention

Electronic Feedback

For each smoking, nutrition, alcohol consumption, and physical activity health risk behavior, if a participant did not meet the Australian health guidelines, information appeared on the computer tablet, specifying that they did not meet the recommended guidelines, and information was provided about effective behavior change strategies. Table 1 presents the Australian health guidelines for each risk behavior and the electronic feedback provided to those who did not meet the recommended guidelines for each behavior.

Table 1. Australian health guidelines and electronic feedback for each smoking, nutrition, alcohol consumption, and physical activity risk behavior.

Health risk behaviors	Australian health guidelines	Electronic feedback
Smoking tobacco	No smoking of cigarettes or tobacco products [47].	• Australian guidelines recommend that not smoking any cigarettes and not using any tobacco improves your health. • Quitting smoking can produce immediate and long-term benefits to your health. • Quitting smoking can take a few attempts so it is important to keep trying. Strategies that can help you quit include web-based and telephone programs, getting your doctor's advice and support, and using nicotine replacement therapies. • We are offering free help to smokers who would like to stop smoking or would consider stopping in the future.
Nutrition (fruit and vegetable intake)	Eat 2 or more serves of fruit each day and 5 or more serves of vegetables each day [48].	• Australian guidelines recommend that adults should eat at least two serves of fruit each day and at least five serves of vegetables each day. Eating healthy can improve your overall health. Strategies that can help you increase your fruit and vegetable intake include web-based and telephone programs, speaking to a dietitian, and getting advice from your doctor. • We are offering free help to people who do not eat enough fruit or vegetables and may like assistance to increase the amount of fruit and vegetables they eat.
Alcohol intake	No more than 2 standard alcoholic drinks per day (to reduce life-time disease risk) and no more than 4 standard alcoholic drinks on one occasion (to reduce the risk of injury and acute problems) [49].	• Australian guidelines recommend that adults should drink no more than 2 standard drinks of alcohol on one day and no more than 4 standard drinks on one occasion to reduce the risk of alcohol-related harm. • Reducing the amount of alcohol you drink can have immediate and long-term benefits to your health. Strategies that can help you reduce your alcohol intake include web-based and telephone programs and getting your doctor's advice and support. • We are offering free help to people who drink more alcohol than recommended and may like assistance to reduce the amount of alcohol they drink.
Physical activity	Do at least 150 to 300 minutes of moderate physical activity or at least 75 to 150 minutes of vigorous physical activity each week [50].	• Australian guidelines recommend that adults do 150 to 300 minutes of moderate physical activity or 75 to 150 minutes of vigorous physical activity each week. Increasing your physical activity can improve your overall health. Strategies that can help you increase your physical activity include web-based and telephone programs, regular times to do physical activity, and participating in sports. • We are offering free help to people who do not do enough physical activity and may like assistance to increase their physical activity.

Proactive Offer of Online and Telephone Services

After the electronic feedback had appeared for each smoking, nutrition, alcohol consumption, and physical activity risk behavior where the participant did not meet Australian health guidelines, both online and telephone services were offered. The online and telephone services were QuitCoach and Quitline for smoking, Healthy Eating Quiz and GHICS for fruit and vegetable consumption, THRIVE and ADIS for risky alcohol consumption, and 10,000 steps and GHICS for physical activity. These support services were chosen because they are existing programs available to the general population in New South Wales for free, and previous studies have demonstrated their effectiveness [27,31,37,51-53]. Participants who agreed to use the online support services were asked to provide their email or mobile phone details. The hyperlinks of the online programs were sent to them. Participants who agreed to sign up for telephone support services were asked to provide their contact details (home and/or mobile phone number) so that the telephone services could contact them.

Measures

Uptake of Proactively Offered Online and Telephone Services

Uptake was measured by participants indicating *yes* they would like to use the specified service and providing the relevant contact details (ie, phone number and/or email address). Uptake referred to signing up to use the services offered.

Sociodemographic Characteristics

Age, sex, country of birth, Aboriginal and Torres Strait Islander identity, highest level of education completed, marital status, employment status, and postcode of residence were collected.

BMI

Participants were asked to self-report their height in centimeters and weight in kilograms. Conversion tables were displayed in the questionnaire to help participants if they required assistance converting their height from feet and inches into centimeters or their weight from stone into kilograms. Participants' BMI scores were calculated and categorized as underweight (BMI<18.5 kg/m^2), healthy weight (BMI 18.5-24.9 kg/m^2), overweight (BMI 25-29.9 kg/m^2), and obese (BMI≥30 kg/m^2).

Anxiety and Depression

The Patient Health Questionnaire-4 (PHQ-4) is a reliable and valid measure that was used to measure anxiety and depression [54]. The PHQ-4 asks, *Over the last 2 weeks, how often have you been bothered by the following problems?* The 2 questions used to measure depression were (1) *Little interest or pleasure in doing things* and (2) *Feeling down, depressed or hopeless*. The 2 items used to measure anxiety were (1) *Feeling nervous, anxious or on edge* and (2) *Not being able to stop or control worrying*. Response options were as follows: not at all (0), several days (1), more than half the days (2), and nearly every day (3) [54]. Scores on both the anxiety and depression scale were dichotomized into yes (ie, those with a total score ≥3 for anxiety or depression) and no (ie, those with a total score <3 for anxiety or depression).

Intention to Change

Participants' intentions to change the behaviors for which they did not meet Australian guidelines were assessed by asking what best described their intentions regarding (1) quitting smoking, (2) reducing alcohol consumption, (3) increasing daily fruit intake, (4) increasing daily vegetable intake, and (5) increasing weekly physical activity. The response options were as follows: will quit (for smoking) or reduce (for alcohol) or increase (for fruit, vegetables, and physical activity) in the next 30 days; will quit (for smoking) or reduce (for alcohol) or increase (for fruit, vegetables, and physical activity) in the next 6 months; may quit (for smoking) or reduce (for alcohol) or increase (for fruit, vegetables, and physical activity) in the future, but not in the next 6 months; never expect to quit (for smoking) or reduce (for alcohol) or increase (for fruit, vegetables, and physical activity); and do not know.

Statistical Analysis

All analyses were performed using SAS version 9.3 (SAS Institute Inc). Frequencies and percentages were used to describe categorical data (ie, sociodemographic characteristics, BMI, and psychological characteristics). Mean and standard deviation were used to describe the continuous data (ie, age of respondents) in the demographics table. The uptake of proactively offered online and telephone support services was described using frequencies and percentages. Four multiple logistic regression models were used to examine whether sociodemographic characteristics, psychological characteristics, and intention to change were associated with signing up for telephone or online services for each smoking, nutrition, alcohol consumption, and physical activity risk behavior. BMI was also included in the regression models for nutrition and physical activity that examined the characteristics associated with signing up for telephone or online support services for these behaviors. A fifth logistic regression model was used to identify whether sociodemographic factors, psychological characteristics, and intention to change were associated with signing up for telephone or online support services for multiple health risk behaviors (ie, at least two risky behaviors).

Results

Participants' Characteristics

Overall, 551 participants had at least one health risk behavior and were offered both online and telephone support services. Table 2 shows the sociodemographic characteristics, BMI, and psychological characteristics of the sample.

Table 2. Sociodemographic, BMI, psychological characteristics, and vocational education courses of the sample (N=551).

Characteristic	Value
Gender, n (%)	
Male	310 (56.3)
Female	228 (41.4)
Other	13 (2.4)
Age, years, mean (SD)	25.7 (11.1)
Country of birth, n (%)	
Australia	509 (92.4)
Other	42 (7.6)
Highest level of education completed, n (%)	
High school or less	388 (70.4)
TAFE[a] or university	163 (29.6)
Marital status, n (%)	
Never married	345 (62.6)
Married or living with partner	166 (30.1)
Divorced or separated or widowed	40 (7.3)
Employment status, n (%)	
Employed	432 (78.4)
Unemployed	119 (21.6)
Aboriginal and Torres Strait Islander, n (%)	
Yes	72 (13.1)
No	479 (86.9)
Residence [b] , n (%)	
Urban	331 (64.4)
Rural	183 (35.6)
BMI status [c] , n (%)	
Underweight	21 (4.0)
Healthy weight	241 (45.6)
Overweight	147 (27.8)
Obese	119 (22.5)
Depression, n (%)	
Yes	131 (23.8)
No	420 (76.2)
Anxiety, n (%)	
Yes	163 (29.6)
No	388 (70.4)
Vocational education course, n (%)	
Automotive mechanic	165 (29.9)
Community service or mental health	56 (10.2)
Floristry	49 (8.9)
Fabrication and welding	38 (6.9)
Hairdressing	37 (6.7)

Characteristic	Value
Mechanical engineering	34 (6.9)
Animal studies or veterinary nursing	33 (6.0)
Arts	23 (4.2)
Dental nursing	22 (4.0)
Electrotechnology	21 (3.8)
Protective coating	19 (3.5)
Programming	15 (2.7)
Baking	13 (2.4)
Mobile plant technology	12 (2.2)
Photography	7 (1.2)
Commercial cookery	5 (0.9)
Business administration	2 (0.4)

[a] TAFE: Technical and Further Education.

[b] Data missing for residence, n=37.

[c] Data missing for BMI, n=23.

Uptake of Proactively Offered Online and Telephone Support Services

Table 3 outlines the uptake of online and telephone support services for each smoking, nutrition, alcohol consumption, and physical activity risk behavior as well as for multiple health risk behaviors. Among those who smoked tobacco, 6.8% (13/191) and 2.6% (5/191) signed up to use QuitCoach and Quitline, respectively, to help them quit smoking, whereas 6.8% (13/191) signed up to use either Quitline or QuitCoach. For participants not meeting the recommended guidelines for fruit and vegetable intake, 14.3% (58/406) and 3.9% (16/406) signed up to use the Healthy Eating Quiz and GHICS, respectively, to improve their fruit and vegetable intake. A total of 14.5% (59/406) signed up for either the Healthy Eating Quiz or GHICS. Of those who were offered online and telephone services to reduce their alcohol consumption, 5.5% (18/327) signed up for THRIVE and 0.9% (3/327) signed up for ADIS, with 5.5% (18/327) signing up for either ADIS or THRIVE. Among those who did not meet the physical activity guidelines, 11.4% (26/228) signed up for 10,000 steps and 7.0% (16/228) signed up for GHICS to improve their physical activity; 12.7% (29/228) signed up for either 10,000 steps or GHICS to modify their physical activity.

Table 3. Uptake of proactively offered online and telephone support services (N=551).

Health risk behaviors and online and telephone interventions	Eligible students offered the support services, n	Signed up to support services, n (%)
Smoking		
Quitline (telephone)	191	5 (2.6)
QuitCoach (online)	191	13 (6.8)
Quitline or QuitCoach	191	13 (6.8)
Inadequate fruit and vegetable consumption		
Get Healthy Information and Coaching Service (telephone)	406	16 (3.9)
Healthy Eating Quiz (online)	406	58 (14.3)
Healthy Eating Quiz or Get Healthy Information and Coaching Service	406	59 (14.5)
Risky alcohol consumption		
ADIS[a] (telephone)	327	3 (0.9)
THRIVE[b] (online)	327	18 (5.5)
THRIVE or ADIS	327	18 (5.5)
Physical inactivity		
Get Healthy Information and Coaching Service (telephone)	228	16 (7.0)
10,000 steps (online)	228	26 (11.4)
10,000 steps or Get Healthy Information and Coaching Service	228	29 (12.7)
Uptake of support services for multiple behaviors		
Uptake of telephone services for multiple behaviors (eg, GHICS[c] and ADIS)	377	8 (2.1)
Uptake of online services for multiple behaviors (eg, THRIVE and Healthy Eating Quiz)	377	21 (5.6)
Uptake of any service (online or telephone) for at least two health behaviors (eg, THRIVE for alcohol and GHICS for physical activity)	377	22 (5.8)

[a] ADIS: Alcohol Drug Information Service.

[b] THRIVE: Tertiary Health Research Intervention Via Email.

[c] GHICS: Get Healthy Information and Coaching Service.

Of the participants who engaged in multiple health risk behaviors, 5.6% (21/377) signed up for online support services for multiple behaviors and 2.1% (8/377) signed up for telephone services for multiple behaviors. For participants with multiple health risk behaviors, 5.8% (22/377) signed up for either telephone or online services for at least two health behaviors.

Characteristics Associated With the Uptake of Proactively Offered Telephone or Online Support Services for Health Risk Behaviors

The characteristics associated with the uptake of proactively offered telephone or online support services for smoking and alcohol consumption (Table 4), nutrition and physical activity (Table 5), and at least two health risk behaviors (Table 6) have been presented.

Table 4. Characteristics associated with uptake of telephone or online support services for smoking and alcohol consumption.

Characteristics	Uptake of QuitCoach or Quitline (smoking)			Uptake of THRIVE[a] or ADIS[b] (alcohol)		
	n (%)	Odds ratio (95% CI)	*P* value	n (%)	Odds ratio (95% CI)	*P* value
Gender			.30			.17
Male	8 (6.4)	Ref[c]		13 (6.2)	Ref	
Female	5 (8.3)	0.33 (0.04-2.65)		5 (4.6)	0.38 (0.10-1.51)	
Age (years)			.64			.49
16-39	11 (6.4)	0.55 (0.04-6.93)		15 (5.2)	0.50 (0.07-3.63)	
>40	2 (18.2)	Ref		3 (13.6)	Ref	.
Highest level of education completed			.72			.63
High school or less	9 (5.8)	0.72 (0.12-4.20)		14 (5.5)	1.41 (0.34-5.80)	
University or TAFE[d]	4 (10.8)	Ref		4 (5.7)	Ref	
Marital status			.77			.42
Never married	8 (6.1)	0.70 (0.14-3.56)		9 (4.1)	0.64 (0.21-1.97)	
Divorced or separated or widowed	2 (13.3)	1.75 (0.13-23.00)		3(15.0)	2.14 (0.32-14.58)	
Married or living with partner	3 (6.7)	Ref		6 (7.0)	Ref	
Employment			.02			.68
Employed	8 (4.8)	0.10 (0.01-0.72)		14 (5.0)	0.72 (0.15-3.37)	
Unemployed	5 (20.0)	Ref		4 (8.2)	Ref	
Residence			.07			.86
Urban	6 (5.8)	0.23 (0.05-1.15)		12 (6.4)	1.10 (0.37-3.29)	
Rural	7 (9.7)	Ref		6 (5.1)	Ref	
Depression			.03			.57
Yes	2 (3.3)	Ref		3 (3.8)	Ref	
No	11(8.4)	10.25 (1.30-80.67)		15 (6.1)	1.53 (0.35-6.75)	
Anxiety			.02			.43
Yes	6 (9.0)	Ref		5 (4.9)	Ref	
No	7 (5.7)	0.11 (0.02-0.71)		13 (5.8)	0.59 (0.16-2.22)	
Aboriginal or Torres Strait Islander			.28			.34
Yes	2 (5.6)	0.34 (0.05-2.39)		1 (2.4)	0.35 (0.04-3.05)	
No	11 (7.1)	Ref		17 (6.0)	Ref	
Intention to change			.08			.13
No intention to change in 6 months or did not know	5 (3.7)	Ref		12 (4.3)	Ref	
Intention to change in 30 days	4 (16.0)	5.46 (1.11-26.97)		4 (13.8)	3.01 (0.76-11.95)	
Intention to change in 6 months	4 (12.5)	3.96 (0.77-20.27)		2 (11.8)	3.63 (0.67-19.62)	

[a] THRIVE: Tertiary Health Research Intervention Via Email.

[b] ADIS: Alcohol and Drug Information Services.

[c] Ref: Reference category

[d] TAFE: Technical and Further Education.

Table 5. Characteristics associated with the uptake of telephone or online support services for nutrition and physical inactivity.

Characteristics	Uptake of Healthy Eating Quiz or GHICS[a] (fruit and vegetables)			Uptake of 10,000 steps or GHICS (physical activity)		
	n (%)	Odds ratio (95% CI)	*P* value	n (%)	Odds ratio (95% CI)	*P* value
Gender			.60			.34
Male	25 (10.9)	Ref[b]		6 (6.1)	Ref	
Female	33 (20.0)	1.22 (0.58-2.56)		22 (17.7)	1.80 (0.54-6.05)	
Age (years)			.78			.72
16-39	48 (14.1)	1.15 (0.42-3.15)		18 (10.0)	0.78 (0.20-3.02)	
>40	11 (22.5)	Ref		11 (25.6)	Ref	
Highest level of education completed			.42			.76
High school or less	34 (11.9)	0.75 (0.38-1.50)		13 (9.1)	0.85 (0.30-2.39)	
University or TAFE[c]	25 (20.7)	Ref		16 (18.8)	Ref	
Marital status			.65			.56
Never married	31 (12.3)	0.79 (0.37-1.67)		13 (10.2)	1.00 (0.31-3.22)	
Divorced or separated or widowed	4 (11.8)	0.60 (0.17-2.11)		5 (25.0)	2.31 (0.49-10.90)	
Married or living with partner	24 (20.0)	Ref		11 (13.7)	Ref	
Employment			.07			.11
Employed	33 (10.5)	0.48 (0.22-1.05)		10 (6.5)	0.42 (0.14-1.24)	
Unemployed	26 (28.6)	Ref		19 (25.3)	Ref	
Residence			.17			.65
Urban	45 (18.5)	1.65 (0.81-3.38)		21 (14.3)	1.29 (0.44-3.80)	
Rural	13 (9.6)	Ref	.	7 (10.3)	Ref	
Depression			.20			.36
Yes	18 (17.7)	Ref		10 (15.9)	Ref	
No	41 (13.5)	0.58 (0.26-1.32)		19 (11.5)	0.57 (0.17-1.90)	
Anxiety			.29			.90
Yes	19 (16.2)	Ref		13 (16.3)	Ref	
No	40 (13.8)	1.58 (0.68-3.64)		16 (10.8)	1.08 (0.32-3.62)	
Aboriginal or Torres Strait Islander			.19			.29
Yes	5 (10.0)	0.43 (0.12-1.52)		3 (10.0)	0.41 (0.08-2.13)	
No	54 (15.2)	Ref		26 (13.1)	Ref	
BMI status			.23			.13
Overweight	11 (10.5)	0.60 (0.25-1.42)		9 (16.1)	3.43 (1.01-11.65)	
Obesity	20 (23.3)	1.39 (0.65-2.97)		12 (21.1)	2.34 (0.68-8.01)	
Healthy weight or underweight	26 (13.2)	Ref		7 (7.1)	Ref	
Intention to change			.25			.04
No intention to change in 6 months or did not know	25 (11.5)	Ref		7 (6.4)	Ref	
Intention to change in 30 days	23 (20.2)	1.76 (0.87-3.56)		15 (24.2)	4.01 (1.33-12.07)	
Intention to change in 6 months	10 (14.1)	1.57 (0.67-3.66)		7 (12.3)	1.57 (0.42-5.86)	

[a] GHICS: Get Healthy Information and Coaching Service.

[b] Ref: Reference category

[c] TAFE: Technical and Further Education.

Table 6. Characteristics associated with the uptake of telephone or online support services for multiple health risk behaviors.

Characteristics	Uptake of any service for multiple health risk behaviors		
	n (%)	Odds ratio (95% CI)	*P* value
Gender			.74
Male	7 (3.2)	Ref[a]	
Female	14 (9.4)	1.22 (0.38-3.86)	
Age (years)			.75
16-39	16 (5.0)	1.25 (0.32-4.94)	
>40	6 (15.4)	Ref	
Highest level of education completed			.69
High school or less	12 (4.4)	0.81 (0.29-2.27)	
University or TAFE[b]	10 (9.9)	Ref	
Marital status			.24
Never married	9 (3.9)	0.47 (0.15-1.45)	
Divorced or separated or widowed	4 (13.3)	1.69 (0.37-7.69)	
Married or living with partner	9 (8.0)	Ref	
Employment			.003
Employed	8 (2.7)	0.18 (0.06-0.56)	
Unemployed	14 (18.2)	Ref	
Residence			.85
Urban	16 (7.2)	0.89 (0.29-2.79)	
Rural	6 (4.7)	Ref	
Depression			.13
Yes	10 (10.2)	Ref	
No	12 (4.3)	0.41 (0.13-1.30)	
Anxiety			.94
Yes	10 (8.5)	Ref	
No	12 (4.6)	1.04 (0.31-3.50)	
Aboriginal or Torres Strait Islander			.54
Yes	3 (5.7)	0.60 (0.12-2.98)	
No	19 (5.9)	Ref	
Intention to change at least two behaviors			.49
No intention to change at least two behaviors	14 (5.1)	Ref	
Intention to change at least two behaviors	8 (8.0)	1.43 (0.52-3.93)	

[a] Ref: Reference category.

[b] TAFE: Technical and Further Education.

Characteristics Associated With Signing Up for QuitCoach or Quitline for Smoking and for THRIVE or ADIS for Alcohol Consumption

After adjusting for the covariates, vocational education students who reported not being depressed had greater odds (odds ratio [OR] 10.25, 95% CI 1.30-80.67) of signing up for QuitCoach or Quitline than those who were depressed. Those who were employed (OR 0.10, 95% CI 0.01-0.72) or reported not being anxious (OR 0.11, 95% CI 0.02-0.71) had smaller odds of signing up for QuitCoach or Quitline.

None of the characteristics were associated with signing up for THRIVE or ADIS to modify alcohol consumption after adjusting for the covariates.

Characteristics Associated With Signing Up for Healthy Eating Quiz or GHICS for Fruit and Vegetable

Intake and for 10,000 Steps or GHICS for Physical Activity

After adjusting for the covariates, none of the factors were associated with signing up for the Healthy Eating Quiz or GHICS for fruit and vegetable consumption.

After adjusting for the covariates, vocational education students who intended to increase their physical activity within 30 days had greater odds (OR 4.01, 95% CI 1.33-12.07) of signing up for 10,000 steps or the GHICS than those who did not intend to change in 6 months or did not know if they intended to change.

Characteristics Associated With Signing Up for Services for Multiple Health Risk Behaviors

Among vocational education students who engaged in multiple health risk behaviors, those who were employed had significantly smaller odds (OR 0.18, 95% CI 0.06-0.56) of signing up for support services for at least two behaviors than those who were unemployed after adjusting for the covariates.

Discussion

Principal Findings

This study examined the uptake of proactively offered online and telephone support services targeting smoking, nutrition, alcohol consumption, and physical activity risk behaviors and multiple health risk behaviors among vocational education students. More than half of the participants were men. Given that most health behavior studies have an overrepresentation of women, this study, in contrast, presents findings where men are well represented and can be reached via TAFE [55]. The findings revealed that the uptake of online and telephone services that targeted smoking, inadequate fruit and vegetable consumption, risky alcohol consumption, and physical inactivity was low among vocational education students. We also found that less than 10% of vocational education students who engaged in multiple health risk behaviors signed up for online or telephone support services to modify at least two health risk behaviors. Vocational education students who were employed and those who did not have symptoms of anxiety were less likely to sign up for support services targeting smoking. Vocational education students who reported no symptoms of depression were more likely to sign up for support services targeting smoking. Vocational education students who intended to change their physical activity in the next 30 days were more likely to sign up for physical activity support services than those who did not intend to change in 6 months or did not know. Finally, vocational education students who were employed were less likely to sign up for support services for at least two health risk behaviors.

Comparison With Prior Work

Uptake of proactively offered telephone support for smoking was relatively low in our study (2.6%) compared with an Australian study that proactively offered telephone support to smokers in the general population (52%) [30] and compared with 82.9% in a study conducted in Hong Kong [45] and 74% in a Denmark study [22]. Differences in the characteristics of the populations in these 3 studies compared with the younger group in this study may provide some explanation for the lower uptake of proactively offered telephone support in vocational education students than in the general population. For example, younger adults may feel that they will not experience health consequences of smoking in the near future and, consequently, may be less likely to wish to quit smoking than their older counterparts. Furthermore, women may be more likely to sign up for support services than men [22,56], and this may in part account for the higher uptake of smoking telephone services reported in previous studies [22,30,45]. Uptake of proactively offered online support (QuitCoach) targeting smoking in our study (6.8%) was similar to that reported by McClure et al [44] in their study (7%). The study by Skov-Ettrup et al [22] reported a higher uptake (69%) of internet intervention for smoking compared with our study. Although the uptake of telephone support for smoking cessation was low (2.6%) among vocational education students, it is comparable with the proportion of smokers in the general population that use the Quitline in Australia (4%) [32] and in the United States (3.5%) [33]. There were no direct age or socioeconomic status comparable data on uptake in these 2 studies. There are a number of potential reasons for the low uptake of smoking services among vocational education students. First, vocational education students may not wish to quit smoking and may believe there are advantages to smoking that override any negative consequences. Second, they may perceive that they are not addicted to cigarettes and feel that they could quit smoking unassisted if they wanted to. Third, they may be part of a social network where their peers smoke tobacco and may feel that they need to continue to smoke to feel socially accepted by their peers. Additional reasons may include their lack of readiness to change their health risk behaviors, perceived inappropriateness of the services they were offered, their preference to change by themselves, and a belief that the support services they were offered would not help them [57].

Compared with the less than 1% of adults in New South Wales who are overweight and obese and who reported using the GHICS [34], our study achieved higher rates (7.0% for physical inactivity and 3.9% for fruit and vegetables) of signing up for this service. However, the rates of vocational education students signing up for this service were lower compared with what was reported in the study by Wolfenden et al [58], in which 23% of New South Wales residents aged 18 years and above in the general community agreed to allow their details to be forwarded to the GHICS when they were proactively called. Furthermore, our study reported a higher (14.3%) rate of signing up for the Healthy Eating Quiz for fruit and vegetables compared with the study where only 0.4% of people aged 16 years and above signed up to use the Healthy Eating Quiz [37]. There is no existing research that has examined the uptake of proactively offered online and telephone support services targeting risky alcohol consumption; therefore, no comparisons can be made with this study.

Although the uptake of support services was low across the health risk behaviors, vocational education students appeared to prefer online support services (5.5%-14.3%) to telephone services (0.9%-7.0%) targeting each behavior and multiple health risk behaviors. This is not surprising given that the TAFE setting is mostly composed of young adults, who are more likely to use the internet [59] and prefer online support services [60]. The convenience of being able to access online programs when they wished to rather than scheduling a particular time to speak with an advisor may also explain why vocational education students preferred online programs to telephone services. In addition, in vocational education students, the uptake of any support services targeting nutrition and physical activity (12.7%-14.5%) appeared higher than that of services for smoking and alcohol (5.5%-6.8%). Possible explanations include that for physical activity and fruit and vegetable consumption, the goal is to increase healthy behaviors, whereas changes to smoking and alcohol consumption involve stopping risky behaviors. Vocational education students may smoke tobacco or drink alcohol to help them cope with stress, depression, or anxiety [61], during social gatherings including work functions with colleagues [62,63], or to feel socially accepted and to improve self-confidence [62,63]. As a result, vocational education students may be less likely to sign up for services to change smoking and alcohol behaviors than nutrition and physical activity behaviors.

Those who were employed had smaller odds of signing up for support services (ie, Quitline or QuitCoach) for smoking and for services for at least two health risk behaviors. This is in contrast with the study by Mak et al [45], which examined the factors that influenced parents who smoke to participate in a proactive telephone intervention for smoking, and the study by Schneider et al [41], which reported that participants who were employed were more likely to participate in an online intervention targeting multiple lifestyle behaviors [41]. Students who worked may have felt that they had less time compared with students who did not work and did not want to sign up for support services they could not commit the time to. We also found that participants who reported not being depressed were more likely to sign up for support services that targeted smoking, whereas those who were not anxious were less likely to sign up for online or telephone support services for smoking. Previous research has found a relationship between smoking cessation and improved mental health [64]. Vocational education students who are smokers and experiencing symptoms of anxiety may thus feel more compelled to sign up for smoking support services not only to quit smoking but also to improve their mental health [64]. Smoking cessation interventions should therefore not only focus on the physical addiction of smoking but also consider the psychological needs of service users to offer coping strategies that may be useful not only for quitting smoking but also for reducing anxiety and psychological distress. Vocational education students who intended to increase their physical activity in 30 days were more likely to sign up for 10,000 steps or the GHICS compared with those who did not intend to increase their physical activity in the next 6 months or did not know if they would change. Intention to change was associated with the uptake of physical activity services but not with the uptake of smoking, nutrition, and alcohol support services. These results suggest that smoking, nutrition, and alcohol support services should be offered to all vocational education students who do not meet the recommended guidelines, irrespective of their intention to change. Physical activity services may benefit from targeting their messaging and recruitment efforts to those intending to increase their physical activity in 30 days. Interventions that target health risk behaviors should therefore consider how to promote their services to vocational education students who do not meet the recommended guidelines for each behavior and how to incorporate strategies to motivate vocational education students to sign up for support services.

Our study offered existing online and telephone services to vocational education students. Although some services such as the GHICS targeted multiple behaviors (ie, fruit and vegetable intake and physical activity), others such as the Quitline focused on one behavior (ie, smoking). There is no telephone service in Australia that addresses all smoking, nutrition, alcohol consumption, and physical activity risk behaviors collectively. Given that smoking, nutrition, alcohol consumption, and physical activity risk behaviors cluster together in vocational education students [6-8] and that transfer theory [10] suggests that modifying one health risk behavior can lead to changes in other behaviors, future studies may wish to offer vocational education students interventions in a form where all smoking, nutrition, alcohol consumption, and physical activity risk behaviors can be addressed by the same service simultaneously.

Limitations

This study had some limitations. First, to be eligible to participate in this study, which was part of a cluster randomized controlled trial, vocational education students needed to be enrolled in a class that ran for at least six months. Therefore, these findings may not be generalizable to vocational education students enrolled in courses that run for less than 6 months. Second, vocational education classes were recruited from the Hunter, Upper Hunter, and Central Coast areas of New South Wales and may not be representative of all vocational education campuses across Australia or internationally.

Conclusions

Although most vocational education students who were offered online and telephone support services did not sign up for these, the uptake rates for some of the support services appear to be higher than self-initiated uptake in the general population. Scaling up the proactive offer of online and telephone services may produce beneficial health outcomes. Proactively offering support services to vocational education students is sustainable via a system whereby students receive electronic feedback about health risk behaviors and referral to existing online and telephone services as part of the standard vocational education enrollment procedure and via student services. The use of existing online and telephone services is an important strength that supports the sustainability of such an intervention in a vocational education setting.

The findings from this study also suggest that vocational education students prefer online support services to telephone services that target smoking, nutrition, alcohol consumption, physical activity and multiple health risk behaviors. Uptake of support services for nutrition and physical activity also appeared higher than that for smoking and alcohol health risk behaviors. This provides important information for developing health interventions for vocational education students in terms of the mode of delivering interventions to vocational education students and the behaviors they prioritize to change. Future studies should also explore what vocational education students perceive to be the barriers that hinder their uptake of online and telephone support services targeting multiple health risk behaviors and what strategies they would be more likely to use.

Acknowledgments

This work was supported by a National Health and Medical Research Council (NHMRC) project grant (APP1124462). PA was supported by a 50/50 Faculty of Health and Medicine, University of Newcastle PhD Scholarship. FT was supported by a Cancer Institute New South Wales Early Career Fellowship (15/ECF/1-44), followed by an NHMRC Career Development Fellowship (APP1143269). The sponsors did not have any influence on the design of the study, data collection, statistical analysis, interpretation of results, and drafting of the manuscript. The authors also want to acknowledge all TAFE students and staff at the participating TAFE campuses for their involvement in this study. Finally, the authors wish to thank Priority Research Centre for Health Behaviour for the grant provided to publish this paper.

Conflicts of Interest

None declared.

References

1. Education at a Glance 2020. Organisation For Economic Co-operation and Development. Paris; 2019. URL: https://www.oecd.org/education/education-at-a-glance/ [accessed 2019-04-16]
2. Bonevski B, Guillaumier A, Paul C, Walsh R. The vocational education setting for health promotion: a survey of students' health risk behaviours and preferences for help. Health Promot J Austr 2013 Dec;24(3):185-191. [doi: 10.1071/HE13047] [Medline: 24355338]
3. Fotheringham MJ, Owen N, Ball K, Clavisi OM, Leslie ER. Campus settings and health promotion: overweight, physical inactivity, smoking and unsafe alcohol consumption in metropolitan university and rural TAFE students. Health Prom J Aust 2000 Dec;10:255-258.
4. Dietz WH, Douglas CE, Brownson RC. Chronic disease prevention: tobacco avoidance, physical activity, and nutrition for a healthy start. J Am Med Assoc 2016 Oct 25;316(16):1645-1646. [doi: 10.1001/jama.2016.14370] [Medline: 27668419]
5. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med 2013 Sep 05;369(10):954-964. [doi: 10.1056/NEJMra1203528] [Medline: 24004122]
6. Bannink R, Broeren S, Heydelberg J, van't Klooster E, Raat H. Depressive symptoms and clustering of risk behaviours among adolescents and young adults attending vocational education: a cross-sectional study. BMC Public Health 2015 Apr 18;15:396 [FREE Full text] [doi: 10.1186/s12889-015-1692-7] [Medline: 25896828]
7. Jeffries JK, Lytle L, Sotres-Alvarez D, Golden S, Aiello AE, Linnan L. Chronic disease risk typologies among young adults in community college. Am J Health Behav 2018 Mar 01;42(2):71-84. [doi: 10.5993/AJHB.42.2.7] [Medline: 29458516]
8. Tomczyk S, Pedersen A, Hanewinkel R, Isensee B, Morgenstern M. Polysubstance use patterns and trajectories in vocational students--a latent transition analysis. Addict Behav 2016 Jul;58:136-141. [doi: 10.1016/j.addbeh.2016.02.027] [Medline: 26943488]
9. Hajian S. Transfer of learning and teaching: a review of transfer theories and effective instructional practices. IAFOR J Edu 2019 Jun 01;7(1). [doi: 10.22492/ije.7.1.06]
10. Lippke S, Nigg CR, Maddock JE. Health-promoting and health-risk behaviors: theory-driven analyses of multiple health behavior change in three international samples. Int J Behav Med 2012 Mar;19(1):1-13 [FREE Full text] [doi: 10.1007/s12529-010-9135-4] [Medline: 21234735]
11. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078. [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
12. Stead LF, Hartmann-Boyce J, Perera R, Lancaster T. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2013;8:CD002850. [doi: 10.1002/14651858.CD002850.pub3] [Medline: 23934971]
13. Goode AD, Reeves MM, Eakin EG. Telephone-delivered interventions for physical activity and dietary behavior change: an updated systematic review. Am J Prev Med 2012 Jan;42(1):81-88. [doi: 10.1016/j.amepre.2011.08.025] [Medline: 22176852]
14. Rose T, Barker M, Maria JC, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
15. Prosser T, Gee KA, Jones F. A meta-analysis of effectiveness of E-interventions to reduce alcohol consumption in college and university students. J Am Coll Health 2018;66(4):292-301. [doi: 10.1080/07448481.2018.1440579] [Medline: 29452058]
16. Meader N, King K, Wright K, Graham HM, Petticrew M, Power C, et al. Multiple risk behavior interventions: meta-analyses of RCTs. Am J Prev Med 2017 Jul;53(1):e19-e30 [FREE Full text] [doi: 10.1016/j.amepre.2017.01.032] [Medline: 28258777]

17. Rochlen AB, Zack JS, Speyer C. Online therapy: review of relevant definitions, debates, and current empirical support. J Clin Psychol 2004 Mar;60(3):269-283. [doi: 10.1002/jclp.10263] [Medline: 14981791]
18. Copeland J, Martin G. Web-based interventions for substance use disorders. J Subst Abuse Treat 2004 Mar;26(2):109-116. [doi: 10.1016/S0740-5472(03)00165-X]
19. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
20. Bywood P, Raven M, Butler C. Telehealth in primary health care settings within Australia and internationally. PHCRIS policy issue review. 2013. URL: http://hdl.handle.net/2328/36229 [accessed 2020-04-20]
21. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]
22. Skov-Ettrup LS, Dalum P, Ekholm O, Tolstrup JS. Reach and uptake of internet- and phone-based smoking cessation interventions: results from a randomized controlled trial. Prev Med 2014 May;62:38-43. [doi: 10.1016/j.ypmed.2014.01.020] [Medline: 24508983]
23. Ybarra ML, Eaton WW. Internet-based mental health interventions. Ment Health Serv Res 2005 Jun;7(2):75-87. [doi: 10.1007/s11020-005-3779-8]
24. Keller PA, Bailey LA, Koss KJ, Baker TB, Fiore MC. Organization, financing, promotion, and cost of US quitlines, 2004. Am J Prev Med 2007 Jan;32(1):32-37. [doi: 10.1016/j.amepre.2006.08.031] [Medline: 17184958]
25. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Outcomes of minimal and moderate support versions of an internet-based diabetes self-management support program. J Gen Intern Med 2010 Dec;25(12):1315-1322 [FREE Full text] [doi: 10.1007/s11606-010-1480-0] [Medline: 20714820]
26. Balmford J, Borland R, Li L, Ferretter I. Usage of an Internet smoking cessation resource: the Australian QuitCoach. Drug Alcohol Rev 2009 Jan;28(1):66-72 [FREE Full text] [doi: 10.1111/j.1465-3362.2008.00009.x] [Medline: 19320678]
27. Borland R, Balmford J, Benda P. Population-level effects of automated smoking cessation help programs: a randomized controlled trial. Addiction 2013 Mar;108(3):618-628. [doi: 10.1111/j.1360-0443.2012.04091.x] [Medline: 22994457]
28. Kypri K, Hallett J, Howat P, McManus A, Maycock B, Bowe S, et al. Randomized controlled trial of proactive web-based alcohol screening and brief intervention for university students. Arch Intern Med 2009 Sep 14;169(16):1508-1514. [doi: 10.1001/archinternmed.2009.249] [Medline: 19752409]
29. Brown WJ, Mummery K, Eakin E, Schofield G. 10,000 Steps Rockhamptonvaluation of a whole community approach to improving population levels of physical activity. J Phys Act Health 2006 Jan;3(1):1-14 [FREE Full text] [doi: 10.1123/jpah.3.1.1]
30. Tzelepis F, Paul CL, Walsh RA, Wiggers J, Knight J, Lecathelinais C, et al. Telephone recruitment into a randomized controlled trial of quitline support. Am J Prev Med 2009 Oct;37(4):324-329. [doi: 10.1016/j.amepre.2009.05.022] [Medline: 19765505]
31. O'Hara BJ, Phongsavan P, Venugopal K, Eakin EG, Eggins D, Caterson H, et al. Effectiveness of Australia's Get Healthy Information and Coaching Service®: translational research with population wide impact. Prev Med 2012 Oct;55(4):292-298. [doi: 10.1016/j.ypmed.2012.07.022] [Medline: 22885323]
32. Miller C. Uptake and effectiveness of the Australian telephone Quitline service in the context of a mass media campaign. Tobacco Control 2003 Sep 01;12(90002):53ii-5358. [doi: 10.1136/tc.12.suppl_2.ii53]
33. Kaufman A, Augustson E, Davis K, Finney Rutten LJ. Awareness and use of tobacco quitlines: evidence from the Health Information National Trends Survey. J Health Commun 2010;15 Suppl 3:264-278. [doi: 10.1080/10810730.2010.526172] [Medline: 21154098]
34. O'Hara BJ, Phongsavan P, Venugopal K, Bauman AE. Characteristics of participants in Australia's Get Healthy telephone-based lifestyle information and coaching service: reaching disadvantaged communities and those most at need. Health Educ Res 2011 Dec;26(6):1097-1106 [FREE Full text] [doi: 10.1093/her/cyr091] [Medline: 21987479]
35. Brown J, Michie S, Raupach T, West R. Prevalence and characteristics of smokers interested in internet-based smoking cessation interventions: cross-sectional findings from a national household survey. J Med Internet Res 2013;15(3):e50 [FREE Full text] [doi: 10.2196/jmir.2342] [Medline: 23506944]
36. Annual Report 2018-19. Australian Bureau of Statistics. 2019 Oct 15. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/1001 [accessed 2020-04-20]
37. Williams RL, Rollo ME, Schumacher T, Collins CE. Diet quality scores of Australian adults who have completed the healthy eating quiz. Nutrients 2017 Aug 15;9(8):- [FREE Full text] [doi: 10.3390/nu9080880] [Medline: 28809805]
38. Prochaska JO, Velicer WF, Fava JL, Rossi JS, Tsoh JY. Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation. Addict Behav 2001;26(4):583-602. [doi: 10.1016/s0306-4603(00)00151-9] [Medline: 11456079]
39. Drehmer JE, Hipple B, Nabi-Burza E, Ossip DJ, Chang Y, Rigotti NA, et al. Proactive enrollment of parents to tobacco quitlines in pediatric practices is associated with greater quitline use: a cross-sectional study. BMC Public Health 2016 Jun 24;16:520 [FREE Full text] [doi: 10.1186/s12889-016-3147-1] [Medline: 27342141]

40. Murray RL, Coleman T, Antoniak M, Stocks J, Fergus A, Britton J, et al. The effect of proactively identifying smokers and offering smoking cessation support in primary care populations: a cluster-randomized trial. Addiction 2008 Jun;103(6):998-1006; discussion 1007. [doi: 10.1111/j.1360-0443.2008.02206.x] [Medline: 18422823]
41. Schneider F, Schulz DN, Pouwels LH, de Vries H, van Osch LA. The use of a proactive dissemination strategy to optimize reach of an internet-delivered computer tailored lifestyle intervention. BMC Public Health 2013 Aug 05;13:721 [FREE Full text] [doi: 10.1186/1471-2458-13-721] [Medline: 23914991]
42. Hammett P, Fu SS, Nelson D, Clothier B, Saul JE, Widome R, et al. A proactive smoking cessation intervention for socioeconomically disadvantaged smokers: the role of smoking-related stigma. Nicotine Tob Res 2018 Feb 07;20(3):286-294 [FREE Full text] [doi: 10.1093/ntr/ntx085] [Medline: 28398492]
43. Wolfenden L, Paul C, Tzelepis F, Freund M, Wiggers J, Gillham K. Acceptability of proactive telephone recruitment to a telephone support service to encourage healthy eating, physical activity and weight loss. Aust N Z J Public Health 2012 Jun;36(3):295-296. [doi: 10.1111/j.1753-6405.2012.00876.x] [Medline: 22672041]
44. McClure JB, Greene SM, Wiese C, Johnson KE, Alexander G, Strecher V. Interest in an online smoking cessation program and effective recruitment strategies: results from Project Quit. J Med Internet Res 2006 Aug 22;8(3):e14 [FREE Full text] [doi: 10.2196/jmir.8.3.e14] [Medline: 16954124]
45. Mak YW, Loke AY, Lam TH, Abdullah AS. Predictors of the participation of smoking parents in a proactive telephone-based smoking cessation program. Addict Behav 2006 Oct;31(10):1731-1743. [doi: 10.1016/j.addbeh.2005.12.018] [Medline: 16448782]
46. Champion KE, Mather M, Spring B, Kay-Lambkin F, Teesson M, Newton NC. Clustering of multiple risk behaviors among a sample of 18-year-old Australians and associations with mental health outcomes: a latent class analysis. Front Public Health 2018;6:135 [FREE Full text] [doi: 10.3389/fpubh.2018.00135] [Medline: 29868543]
47. National Drug Strategy Household Survey 2016: detailed findings. Australian Institute of Health and Welfare. Canberra; 2017. URL: https://www.aihw.gov.au/getmedia/15db8c15-7062-4cde-bfa4-3c2079f30af3/21028a.pdf.aspx?inline=true [accessed 2019-06-22]
48. Australian Dietary Guidelines. National Health and Medical Research Council. Canberra; 2013. URL: https://www.health.gov.au/sites/default/files/australian-dietary-guidelines.pdf [accessed 2018-07-10]
49. Australian Guidelines to Reduce Health Risks from Drinking Alcohol. National Health and Medical Research Council. Canberra: National Health and Medical Research Council; 2009. URL: https://www.nhmrc.gov.au/about-us/publications/australian-guidelines-reduce-health-risks-drinking-alcohol#block-views-block-file-attachments-content-block-1 [accessed 2019-08-14]
50. Physical activity across the life stages. Australian Institute of Health and Welfare. Canberra: AIHW; 2018. URL: https://www.aihw.gov.au/getmedia/c249ef97-e219-44df-a8bd-f5e50d04064c/aihw-phe-225.pdf.aspx?inline=true [accessed 2019-05-20]
51. Kwan MY, Cairney J, Faulkner GE, Pullenayegum EE. Physical activity and other health-risk behaviors during the transition into early adulthood: a longitudinal cohort study. Am J Prev Med 2012 Jan;42(1):14-20. [doi: 10.1016/j.amepre.2011.08.026] [Medline: 22176841]
52. Davies C, Corry K, Van IA, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
53. Tzelepis F, Paul CL, Wiggers J, Walsh RA, Knight J, Duncan SL, et al. A randomised controlled trial of proactive telephone counselling on cold-called smokers' cessation rates. Tob Control 2011 Jan;20(1):40-46 [FREE Full text] [doi: 10.1136/tc.2010.035956] [Medline: 21030529]
54. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]
55. Maher CA, Lewis LK, Ferrar K, Marshall S, De BI, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
56. Zbikowski SM, Hapgood J, Smucker BS, McAfee T. Phone and web-based tobacco cessation treatment: real-world utilization patterns and outcomes for 11,000 tobacco users. J Med Internet Res 2008;10(5):e41 [FREE Full text] [doi: 10.2196/jmir.999] [Medline: 19017583]
57. Paul CL, Wiggers J, Daly JB, Green S, Walsh RA, Knight J, et al. Direct telemarketing of smoking cessation interventions: will smokers take the call? Addiction 2004 Jul;99(7):907-913. [doi: 10.1111/j.1360-0443.2004.00773.x] [Medline: 15200586]
58. Wolfenden L, Wiggers J, Paul C, Freund M, Lecathelinais C, Wye P, et al. Increasing the use of preventative health services to promote healthy eating, physical activity and weight management: the acceptability and potential effectiveness of a proactive telemarketing approach. BMC Public Health 2012 Nov 07;12:953 [FREE Full text] [doi: 10.1186/1471-2458-12-953] [Medline: 23134686]

59. Taylor K, Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center. Washington: Pew Research Center; 2019 Feb. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-03-20]
60. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
61. Deasy C, Coughlan B, Pironom J, Jourdan D, Mannix-McNamara P. Psychological distress and coping amongst higher education students: a mixed method enquiry. PLoS One 2014;9(12):e115193 [FREE Full text] [doi: 10.1371/journal.pone.0115193] [Medline: 25506825]
62. Reid M, Farrelly F, Farrell L, Fry T, Worsley T. Drinking Related Lifestyles: exploring the Role of Alcohol in Victorians' Lives. Victorian Health Promotion Foundation. Melbourne, Australia; 2013. URL: https://www.vichealth.vic.gov.au/-/media/ResourceCentre/PublicationsandResources/alcohol-misuse/Drinking-Related-Lifestyles/RMIT-DRL_Qualitative-Report.pdf?la=en&hash=B8CFBD894F0F6E05245D1937CA0B0E426BA34CD5 [accessed 2020-04-16]
63. Song AV, Ling PM. Social smoking among young adults: investigation of intentions and attempts to quit. Am J Public Health 2011 Jul;101(7):1291-1296 [FREE Full text] [doi: 10.2105/AJPH.2010.300012] [Medline: 21566040]
64. Taylor G, McNeill A, Girling A, Farley A, Lindson-Hawley N, Aveyard P. Change in mental health after smoking cessation: systematic review and meta-analysis. Br Med J 2014 Feb 13;348:g1151 [FREE Full text] [doi: 10.1136/bmj.g1151] [Medline: 24524926]

Abbreviations

ADIS: Alcohol and Drug Information Services
GHICS: Get Healthy Information and Coaching Service
NHMRC: National Health and Medical Research Council
OR: odds ratio
PHQ-4: Patient Health Questionnaire-4
TAFE: technical and further education
THRIVE: Tertiary Health Research Intervention Via Email

Edited by G Eysenbach, R Kukafka; submitted 05.05.20; peer-reviewed by K Champion, A Leppänen, B Parker; comments to author 12.06.20; revised version received 22.07.20; accepted 11.11.20; published 06.01.21.

Please cite as:
Atorkey P, Paul C, Bonevski B, Wiggers J, Mitchell A, Byrnes E, Lecathelinais C, Tzelepis F
Uptake of Proactively Offered Online and Telephone Support Services Targeting Multiple Health Risk Behaviors Among Vocational Education Students: Process Evaluation of a Cluster Randomized Controlled Trial
J Med Internet Res 2021;23(1):e19737
URL: https://www.jmir.org/2021/1/e19737
doi:10.2196/19737
PMID:33404504

Original Paper

Resource-Based Internet Intervention (Med-Stress) to Improve Well-Being Among Medical Professionals: Randomized Controlled Trial

Ewelina Smoktunowicz[1,2], PhD; Magdalena Lesnierowska[1], PhD; Per Carlbring[2], PhD; Gerhard Andersson[3,4], PhD; Roman Cieslak[1,5], PhD

[1]Department of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland
[2]Department of Psychology, Stockholm University, Stockholm, Sweden
[3]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden
[4]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[5]Trauma, Health, and Hazards Center, University of Colorado Colorado Springs, Colorado Springs, CO, United States

Corresponding Author:
Ewelina Smoktunowicz, PhD
Department of Psychology
SWPS University of Social Sciences and Humanities
Warsaw
Poland
Phone: 48 506354565
Email: esmoktunowicz@swps.edu.pl

Abstract

Background: Medical professionals are exposed to multiple and often excessive demands in their work environment. Low-intensity internet interventions allow them to benefit from psychological support even when institutional help is not available. Focusing on enhancing psychological resources—self-efficacy and perceived social support—makes an intervention relevant for various occupations within the medical profession. Previously, these resources were found to operate both individually or sequentially with self-efficacy either preceding social support (cultivation process) or following it (enabling process).

Objective: The objective of this randomized controlled trial is to compare the efficacy of 4 variants of Med-Stress, a self-guided internet intervention that aims to improve the multifaceted well-being of medical professionals.

Methods: This study was conducted before the COVID-19 pandemic. Participants (N=1240) were recruited mainly via media campaigns and social media targeted ads. They were assigned to 1 of the following 4 groups: experimental condition reflecting the cultivation process, experimental condition reflecting the enabling process, active comparator enhancing only self-efficacy, and active comparator enhancing only perceived social support. Outcomes included 5 facets of well-being: job stress, job burnout, work engagement, depression, and job-related traumatic stress. Measurements were taken on the web at baseline (time 1), immediately after intervention (time 2), and at a 6-month follow-up (time 3). To analyze the data, linear mixed effects models were used on the intention-to-treat sample. The trial was partially blinded as the information about the duration of the trial, which was different for experimental and control conditions, was public.

Results: At time 2, job stress was lower in the condition reflecting the cultivation process than in the one enhancing social support only (d=−0.21), and at time 3, participants in that experimental condition reported the lowest job stress when compared with all 3 remaining study groups (*ds* between −0.24 and −0.41). For job-related traumatic stress, we found a significant difference between study groups only at time 3: stress was lower in the experimental condition in which self-efficacy was enhanced first than in the active comparator enhancing solely social support (d=−0.24). The same result was found for work engagement (d=−0.20), which means that it was lower in exactly the same condition that was found beneficial for stress relief. There were no differences between study conditions for burnout and depression neither at time 2 nor at time 3. There was a high dropout in the study (1023/1240, 82.50% at posttest), reflecting the pragmatic nature of this trial.

Conclusions: The Med-Stress internet intervention improves some components of well-being—most notably job stress—when activities are completed in a specific sequence. The decrease in work engagement could support the notion of dark side of this phenomenon, but further research is needed.

Trial Registration: ClinicalTrials.gov NCT03475290; https://clinicaltrials.gov/ct2/show/NCT03475290

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-019-3401-9

(J Med Internet Res 2021;23(1):e21445) doi:10.2196/21445

KEYWORDS

well-being; job stress; internet intervention; self-efficacy; perceived social support; medical professionals

Introduction

Background

Well-being at work has been the focus of numerous workplace interventions, as it is associated with important outcomes both for the organization and the individual, including performance [1], absenteeism [2], and turnover intentions [3]. One occupational group that is particularly at risk of losses in terms of well-being is health care professionals [4-6]. In this study, we tested the efficacy of an internet intervention that aims to enhance psychological resources—perceived social support and self-efficacy—to improve the well-being of medical professionals. As this study was conducted before the outbreak of the COVID-19 pandemic in 2020, we managed to capture the everyday functioning of medical professionals.

Regardless of occupation, workplace interventions represent various approaches. Some of them are contextual and focus on factors (job demands and resources) that are specific to a given organization [7,8], whereas others are context-free because they target personal characteristics, such as self-efficacy, hope, and optimism [8]. Finally, there are psychotherapies [9,10] that focus on specific mental health disorders, such as depression and anxiety, albeit in the context of work.

Although many of these interventions have been found to be effective, their accessibility is usually constrained, and they can be difficult and costly to scale-up. Technology overcomes some of these limitations. Over the past three decades, internet interventions have been shown to deliver effective treatments for a variety of mental health problems [11-13]. More recently, their efficacy has also been tested in work environments with promising results [14,15]. Some of these interventions have been conceptualized within the cognitive behavioral therapy (CBT) framework [16], whereas others employed mindfulness [17] or stress management techniques [18]. However, although some programs [19] were based on theoretical models such as the stress and cognitive appraisal model by Lazarus and Folkman [20], they rarely test frameworks that have been specifically developed in the context of occupational health (eg, job demands-resources model [21]) or adapted to it (eg, conservation of resources theory [COR] [22]).

COR [22] originated as a conceptual framework of general stress but has since repeatedly served as a backbone for research on occupational well-being [23,24]. One of its fundamental premises is the resource investment principle, also known as a gain spiral. It refers to a process in which people who possess certain resources (eg, time) invest them to generate new ones (eg, learn new skills) and ultimately lower their risk of future losses. This process has already been empirically demonstrated [25-27]; however, studies have tended to focus on contextual job resources and have rarely been experimental (with notable exceptions [28]). Meanwhile, personal resources are by definition context-free. A risk of focusing on them in an intervention is that they are difficult to customize to the unique combination of demands and resources in a given workplace. However, at the same time, such an intervention has the advantage of not being limited to only a few workplaces but is applicable across organizations.

Resources that have repeatedly been shown to be associated with well-being at work are social support and self-efficacy. Social support is usually defined in relation to its source (eg, from supervisor, friends, or family), and within those relationships, it can be either received or provided [29]. Social support that does not refer to the outside source is a perceived one and is defined as one’s belief that help will be provided when needed [30], making it similar to optimism and hope [31]. Moreover, the relationship between received and perceived support is only moderate [31], showing that they are distinct concepts. Therefore, it is justified to enhance perceived social support independently from a specific social context. Finally, social support is a process that generates new resources [32], which makes it uniquely suitable as a chain in the gain spiral. Meanwhile, self-efficacy, another key resource, is defined as “beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” [33]. Its 4 major sources have been identified as mastery and vicarious experiences, persuasion, and reinterpretation of emotional and physiological states [33], with the first considered the most powerful [34]. As self-efficacy is more predictive of a given outcome when it is task- or domain-specific [35], it is more beneficial to enhance contextual self-efficacy in a workplace intervention (eg, related to coping with job stress) than a general one.

Hobfoll et al [32] suggested that people abundant in personal resources such as self-efficacy are in a better position to elicit support from others but also that this relationship works in reverse, and social support can prevent the depletion of personal resources. Called the cultivation and enabling hypotheses, this two-way relationship was later conceptualized in the form of 2 arguments [36]. According to the former, people with higher self-efficacy are more likely to reach out and solicit social support when needed, whereas the latter suggests that using members of a social network as models—in particular those that are similar and/or face similar situations—or benefiting from their verbal assurance can increase the sense of efficacy. Not only theoretical but also empirical support was found for both propositions, however mostly in the context of health [37] and traumatic stress [38,39] but not for outcomes specific to work. In addition, although many of the studies were longitudinal, few were of experimental design. Therefore, in this study, we aim to verify the efficacy of enhancing social support and self-efficacy to improve well-being in a randomized controlled trial (RCT) and test whether the sequence in which they are strengthened is important.

The results of the sixth European Working Conditions Survey show that people working in the health sector experience the highest level of work intensity [40]. Their well-being impacts the services they provide as well as the functioning of health care systems [41]. Thus, improving it is important not only for themselves but also for the patients and their entire workplace communities.

The term well-being covers numerous occupational phenomena. In this intervention, we focus on 5 that are prevalent in the health care setting: job stress, job burnout, work engagement, job-related traumatic stress, and depression. Our primary focus is on job stress and burnout. Stress is a response to a situation when resources are not available or are insufficient to offset demands that are placed upon employees [42], whereas job burnout, which has recently been added to the 11th Revision of the International Classification of Diseases [43], is defined as “a prolonged response to chronic emotional and interpersonal stressors on the job” [44]. Depending on the conceptualization, it comprises 1, 2, or 3 dimensions [45]. Although it is difficult to assess the prevalence of occupational issues such as job stress and job burnout among medical professionals, there is evidence to show that these problems are encountered globally, including in the cultural West [5,46], sub-Saharan Africa [47], and Japan [48]. A recent systematic review that extracted data from populations of 45 countries reported that up to 80.5% of physicians experienced burnout [49], making this a problem that needs to be addressed urgently.

Work engagement is defined as “a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption” [50]. Although it is negatively related to job burnout, they are not opposite poles of one dimension. Recent findings show that employees exhibit different burnout-engagement profiles: from those representing clear discrepancies (ie, high engagement with low burnout and vice versa) to profiles showing that people can be low, moderate, or high on both these dimensions at the same time [51]. Thus, interventions that target job burnout should also measure their impact on work engagement.

Job-related traumatic stress is another outcome that belongs to the occupational context, albeit not to all professions. Defined as stress resulting particularly from indirect exposure to aversive details of traumatic events via face-to-face contact with traumatized individuals or exposure to drastic materials [52], it has been widely recognized as a significant occupational burden among health professionals that needs addressing [53,54].

Contrary to previous outcomes, depression is not a typical occupational phenomenon. However, it is an important component of work-related well-being: it is prevalent across professions and has been found to predict poor job performance [55] and sickness absences [56]. Moreover, although criteria distinguishing depression from job burnout are defined, it remains problematic to differentiate them [57].

Objectives

The main goal of this study was to test whether a Med-Stress internet intervention dedicated to medical professionals would be effective in reducing job stress, burnout, depression, and job-related secondary stress and in increasing work engagement through the enhancement of 2 psychological resources: perceived social support and self-efficacy. We compared 2 experimental conditions, in which self-efficacy and social support were enhanced sequentially, with 2 active controls, in which only self-efficacy or only social support was targeted. We expected that experimental conditions, which are designed to build 2 resources, would prove to be more beneficial than active control ones—that target only a single resource—in terms of reducing negative outcomes (ie, job stress, burnout, depression, and job-related secondary stress) and improving a positive one (ie, work engagement) both at posttest and at a 6-month follow-up. Furthermore, we aim to test whether one sequence of resources—in line with either the enabling or cultivation hypothesis—would turn out to be more beneficial than the other.

Methods

Study Design

The Med-Stress study was designed as a 4-arm parallel RCT comparing the posttest and follow-up effects of 2 experimental and 2 active control conditions. The trial was conducted on web. This study was approved by the Ethical Review Board at SWPS University of Social Sciences and Humanities (opinion 4/2018 issued on February 13, 2018) and registered on ClinicalTrials.gov (NCT03475290). The study protocol has been previously published [58].

Recruitment and Study Participants

Participants were recruited between October 2018 and April 2019 via targeted social media campaigns, advertisements via radio and press, professional forums, and through medical organizations. The following inclusion criteria were applied: (1) being at least 18 years old and (2) representing the health-related profession that involved direct patient care. A total of 1575 people were registered for the study; however, 335 did not complete the required baseline assessment. The final sample size was N=1240.

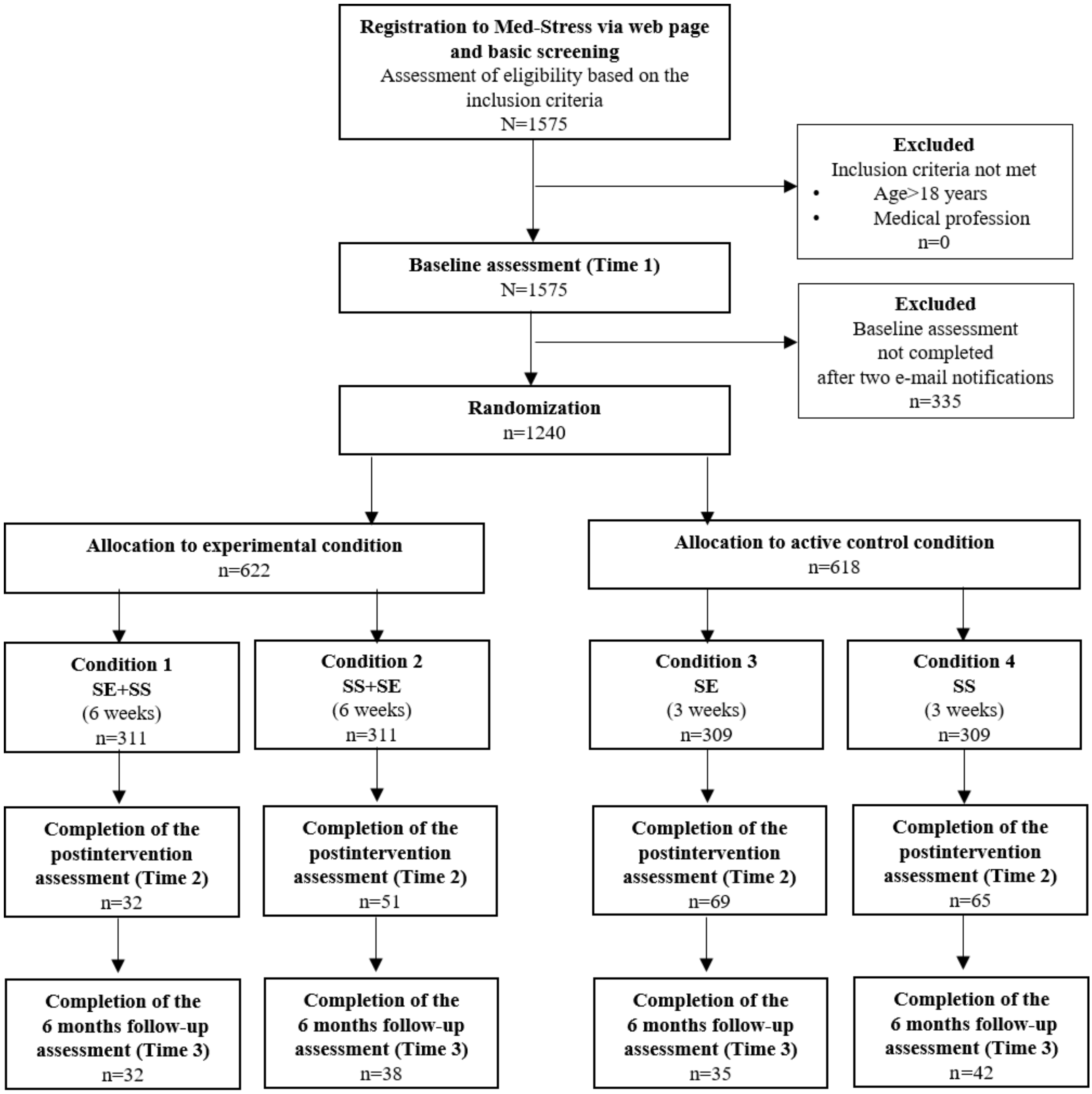
Power Calculation

To ensure a statistical power of 0.90 to detect the posttest effect of comparisons between study conditions, we conducted an a priori sample size estimation using G*Power 3.1 3.1 [59]. Taking into account the previous studies on the effectiveness of internet interventions for general and job-related stress [15,60], we aimed to detect the minimum effect sizes of d=0.40 for the comparisons between conditions at 2 measurement points while controlling for baseline scores at an alpha error level of .002, reduced to correct for multiple comparisons (ie, 5 between-group comparisons and 5 outcomes). A power analysis showed that a sample of 607 was needed. As we expected a high dropout rate, we aimed to recruit a sample of 1200. As the analyses were ultimately conducted using linear mixed effects models (*see the Statistical Analysis* section for more details), we conducted an additional analysis using the powerlmm package in R [61] to find that with the sample of 1240 participants, we had the power to detect an effect size of approximately *d*=0.21 in pairwise comparisons. The dropout rate was higher than expected (1023/1240, 82.50%), and thus, we did not have enough power to test for robustness of effects in the per protocol analysis.

Procedure

The study flow is presented in Figure 1. Participants were directed to the Med-Stress website, where they filled out a screening to ensure they met the inclusion criteria. The registration process was finalized when they signed a web-based informed consent form. Subsequently, participants were asked to fill out a baseline (time 1) assessment. Only those who completed this step were randomized to 1 of the 4 conditions. The intervention lasted either 6 weeks (experimental conditions) or 3 weeks (active controls). Activities in the intervention were released once a week for a participant. Although we encouraged participants to complete all tasks, this was not a prerequisite to access subsequent exercises. Immediately after completing the intervention (at the posttest, time 2) and 6 months after the baseline assessment (follow-up, time 3), participants were invited to fill out the questionnaires on web. Each time, participants received 3 email remainders (2 automatic and 1 personalized).

Figure 1. Flow of participants. SE: self-efficacy enhancement module; SE+SS: self-efficacy and perceived social support sequential enhancement module; SS: perceived social support enhancement module; SS+SE: perceived social support and self-efficacy sequential enhancement module.



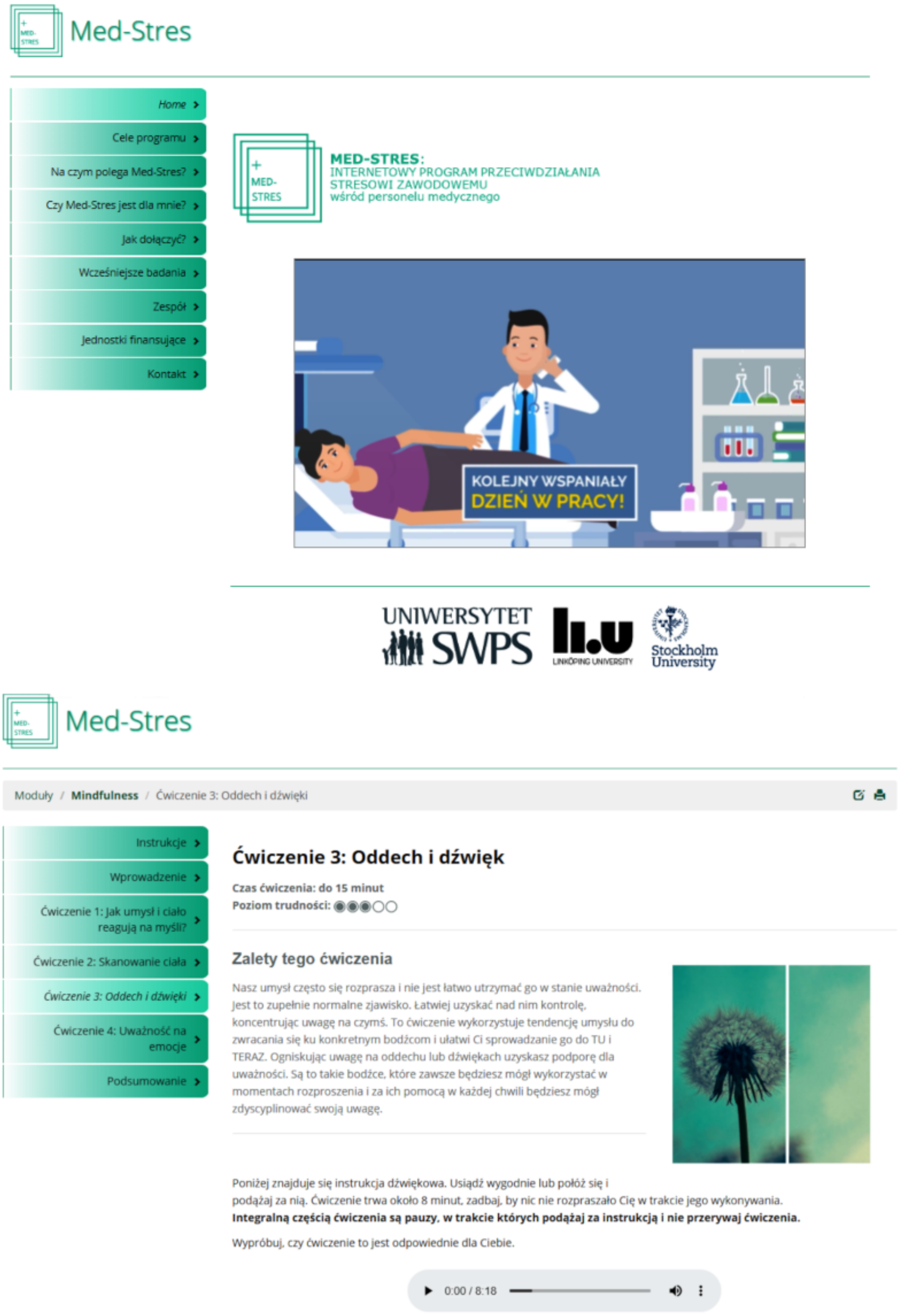
Intervention

Med-Stress is a self-guided internet intervention. It contains 2 main modules that were made available to participants in different variants depending on randomization to study conditions: (1) self-efficacy and perceived social support sequential enhancement modules (SE+SS; experimental condition), (2) perceived social support and self-efficacy sequential enhancement modules (SS+SE; experimental condition), (3) self-efficacy enhancement module (SE; active control condition), and (4) perceived social support enhancement module (SS; active control condition). There were 3 exercises per module. Each consisted of psychoeducational animated clips and interactive tasks requiring both web-based and offline activities. To complete all tasks within each exercise, participants needed up to 1.5 hours. In addition, everyone had access to 4 modules: relaxation, mindfulness, cognitive restructuring, and lifestyle, which were optional and available throughout the intervention. Such a design allowed the participants to partially self-tailor the intervention. Exercises in each module were evidence based and developed in a participatory manner, as a result of the preimplementation study among 744 medical professionals conducted at the Med-Stress developmental stage (Lesnierowska et al, unpublished data, 2018). Exercise descriptions are presented in Table 1, and the example screenshots are shown in Figure 2. The program was run on the Iterapi platform developed at the Department of Behavioral Sciences and Learning at Linköping University, Sweden [62].

Table 1. The content of the Med-Stress intervention: obligatory and optional modules.

Modules	Exercises	Aims
Obligatory		
Self-efficacy enhancement	1. Mastery experience 2. Vicarious experience 3. Action planning	To enhance self-efficacy via recollecting past successes in coping with job stress, taking inspiration from other people (models), and creating a coping plan, based on the "if-then" approach to address potential coping barriers. The exercises are released weekly.
Perceived social support enhancement	1. Received support and cognitive distortions 2. Social skills and peer support 3. Action planning	To enhance perceived social support through verifying distorted assumptions about obtaining support, recalling past situations when the actual support was received, engaging into real receive-provide support interactions with other participants (chat-rooms), and creating a coping plan, based on the "if-then" approach aimed at obtaining social support. The exercises are released weekly.
Optional		
Relaxation	1. Breathing 2. Progressive relaxation 3. Visualizing the body's warmth and weight 4. Visualizing a calm place	To support the coping process by stress relief techniques and actions. The exercises in this module aim at active relaxation. The exercises are available during the whole intervention.
Mindfulness	1. How do body and mind react to thoughts? 2. Body scanning 3. Breathing and sounds 4. Being mindful of emotions	To support the coping process by stress relief techniques and actions. The exercises in this module aim at deliberate attention on a present experience. The exercises are available during the whole intervention
Cognitive restructuring	1. Opinion or fact 2. Identifying thinking traps 3. How important will this be in the future?	To support the coping process by stress relief techniques and actions. The exercises in this module aim at skills such as distinguishing opinions from facts or identifying cognitive traps. The exercises are available during the whole intervention
Lifestyle	1. Physical activity 2. Pleasant activities	To support the coping process by stress relief techniques and actions. The exercises in this module aim at implementing physical activity or pleasant activities that foster stress reduction. The exercises are available during the whole intervention

Figure 2. Med-Stress internet intervention: (a) mastery experience and (b) mindfulness exercises.



MED-STRES:
INTERNETOWY PROGRAM PRZECIWDZIAŁANIA
STRESOWI ZAWODOWEMU
wśród personelu medycznego

Ćwiczenie 3: Oddech i dźwięk

Czas ćwiczenia: do 15 minut
Poziom trudności: ●●●○○

Zalety tego ćwiczenia

Nasz umysł często się rozprasza i nie jest łatwo utrzymać go w stanie uważności. Jest to zupełnie normalne zjawisko. Łatwiej uzyskać nad nim kontrolę, koncentrując uwagę na czymś. To ćwiczenie wykorzystuje tendencję umysłu do zwracania się ku konkretnym bodźcom i ułatwi Ci sprowadzanie go do TU i TERAZ. Ogniskując uwagę na oddechu lub dźwiękach uzyskasz podporę dla uważności. Są to takie bodźce, które zawsze będziesz mógł wykorzystać w momentach rozproszenia i za ich pomocą w każdej chwili będziesz mógł zdyscyplinować swoją uwagę.

Poniżej znajduje się instrukcja dźwiękowa. Usiądź wygodnie lub połóż się i podążaj za nią. Ćwiczenie trwa około 8 minut, zadbaj, by nic nie rozpraszało Cię w trakcie jego wykonywania. **Integralną częścią ćwiczenia są pauzy, w trakcie których podążaj za instrukcją i nie przerywaj ćwiczenia.**

Wypróbuj, czy ćwiczenie to jest odpowiednie dla Ciebie.

Randomization Procedure

Participants were automatically allotted to 1 of the 4 study conditions in line with the predefined randomization protocol generated via randomizing software (randomizer.org, a free service offered by Social Psychology Network). We applied block randomization with a block size of 4. The allocation ratio was 1:1:1:1 to ensure an equal number of participants in each study condition. Masking was partially achieved: allocation was not revealed to the participants; however, they had previously been informed that the intervention would last 3 or 6 weeks depending on the intervention variant.

Measures

Primary Outcomes

Job Stress

Job stress was assessed using the 14-item Perceived Stress Scale [42]. Instructions were adapted for the occupational context. The questionnaire includes 14 items describing the range of stress symptoms rated from 0 (never) to 4 (very often). Cronbach alphas were .85 at time 1, .86 at time 2, and .89 at time 3.

Job Burnout

Job burnout was measured using the Oldenburg Burnout Inventory [63]. The scale consists of 16 items. Participants were asked to respond on a scale ranging from 1 (strongly agree) to 4 (strongly disagree). The reliability of the scale was α=.87 at time 1, α=.89 at time 2, and α=.91 at time 3.

Secondary Outcomes

Depression

Depression was assessed using the Patient Health Questionnaire-9 [64]. The scale has 9 items with responses ranging from 0 (not at all) to 3 (nearly every day). Cronbach alphas were α=.87 at time 1, α=.88 at time 2, and α=.89 at time 3.

Work Engagement

Work engagement was measured using the Utrecht Work Engagement Scale-3 [65]. This shortened tool includes 3 items with a response scale ranging from 0 (never) to 6 (always). The Cronbach alphas were .73 at time 1, .78 at time 2, and .83 at time 3.

Job-Related Traumatic Stress

Job-related traumatic stress was assessed using the Posttraumatic Stress Disorder Checklist 5 [66] with modified instructions. It consists of 20 items. Participants assessed the severity of the symptoms on a scale ranging from 0 (not at all) to 4 (extremely). The Cronbach alphas were α=.95 at time 1, α=.96 at time 2, and α=.96 at time 3.

Other Measures

Self-Efficacy to Manage Job Burnout and Work Stress

Self-efficacy to manage job burnout and work stress was measured using the Work Stress and Job Burnout Self-Efficacy Scale [67]. The scale has 28 items rated from 1 (I am definitely not capable) to 7 (I am definitely capable). The Cronbach alphas were .93 at time 1 and .96 at both time 2 and time 3.

Social Support Self-Efficacy

Social support self-efficacy was measured using the modified version of the Berlin Social Support Scales, Subscale 3 [68]. The tool was adapted to the context of social support self-efficacy. It consists of 5 items with a response scale ranging from 1 (I am definitely not capable) to 7 (I am definitely capable). The reliability was α=.88 at time 1, α=.92 at time 2, and α=.83 at time 3.

Perceived Social Support

Perceived social support was assessed using the Who Can You Count On Scale [69]. The questionnaire consists of 32 items with a response scale ranging from 1 (to very little extent) to 5 (to a great extent). The Cronbach alphas were .95 at time 1 and .96 at time 2 and time 3.

Secondary Trauma Exposure

Secondary trauma exposure was measured using The Secondary Trauma Exposure Scale [70]. Participants indicated their answers on a yes or no scale in response to whether they indirectly experienced each of the 10 listed traumatic events. The volume, frequency, and ratio of exposure were assessed using a single item with response scales ranging from 1 (none or never) to 7 (a few thousands or every day), and from 0% to 100% for the exposure ratio. Due to scale construction, the Cronbach coefficient was not computed.

Expectancy of Improvement and Perceived Credibility of the Intervention

Expectancy of improvement and perceived credibility of the intervention were assessed using the Credibility and Expectancy Questionnaire Version II [71]. The scale consists of 6 items representing 2 subscales: perceived intervention credibility and expectancy of improvement. Responses ranged from 1 (not at all) to 9 (very). The reliability for the perceived intervention credibility subscale was α=.81 and for the expectancy of improvement was α=.89.

Intervention Usage

We assessed the use of obligatory and optional modules. For the former, engaging in at least one weekly assigned task was coded as activity. Optional modules contained a number of similar exercises that participants could choose from, thus engaging in at least one of them was coded as completion.

Statistical Analysis

Analysis of variance (ANOVA) for continuous and chi-square tests for categorical data were used to conduct a randomization check and dropout analysis. We did not formulate predictions regarding differences in resources (ie, perceived social support and self-efficacy) between study conditions, but we did expect them to increase between time 1 and time 2 and to remain stable by time 3. To test the change over time, we applied linear mixed effects models with a restricted maximum likelihood estimation, an unstructured covariance matrix, and time as a fixed factor compared between 3 measurement points.

To verify the effect of intervention conditions on the 5 outcome variables over the course of 3 time points, we applied linear mixed effects models with restricted maximum likelihood estimation, unstructured covariance matrices of random effects, and random slopes and intercepts [72], separately for each of the 5 outcomes. Time variable was coded as time 1=0 (baseline), time 2=1 (posttest), and time 3=2 (follow-up). To make comparisons between each active control group and experimental group as well as between 2 experimental conditions, we created 3 dummy variables in which 2 controls and 1 experimental condition subsequently served as a reference category. It is important to note that our research question was not about the differences in changes over time between groups but about the differences in the outcomes between study conditions at 2 time points: the posttest and at the follow-up. This was reflected in the study protocol [58] where we planned to conduct separate analyses at 2 measurement points. Ultimately, to take into account the dependency between multiple measurements, we used linear mixed effects models. For each outcome, we applied a top-down approach by first fitting a loaded model with fixed and random effects, and we subsequently compared it with 2 nested models: one without random slope and one without random intercept. In the case of job-related secondary traumatic stress, we additionally controlled for exposure to traumatic events, which is a critical factor for the development of posttraumatic symptoms [52]. Finally, we tested whether there were predictors of the 2 primary outcomes (job stress and burnout). To do so, we first calculated the residual gains by subtracting the standardized initial score (Z1) multiplied by the correlation between the initial and posttreatment scores from the standardized posttreatment score (Z2) [73]. Next, we inverted the gains' values so that positive scores reflected reduction. These residual gains were entered separately as outcome variables in the regression analyses performed with 10,000 bootstrapped replications. All analyses in this study were conducted on the intention-to-treat sample. The multiple imputation (MI) procedure was applied using 10 iterations. All baseline scores, available posttest and follow-up scores, and all variables that were found to differentiate completers and dropouts (see below in *Preliminary Results*) were introduced as predictors of imputation [74]. Ultimately, the total sample of 1240 participants was included in the analyses (311 participants in each experimental condition and 309 participants in each control condition; Figure 1). Analyses were conducted using IBM SPSS Statistics 25.0.

Results

Sample

The final sample (N=1240) consisted of physicians (448/1240, 36.13%), nurses (308/1240, 24.84%), dentists (93/1240, 7.50%), physiotherapists (75/1240, 6.05%), midwives (65/1240, 5.24%), paramedics (62/1240, 5.00%), and other medical professionals (191/1240, 15.40%). Overall, job tenure ranged from less than 1 year to 40 years (mean 11.46, SD 10.26), whereas tenure at participants' current position was mean 6.69 years (SD 7.31). On average, participants worked 47.38 hours per week (SD 18.53). The majority of responders were employed in public health care institutions (924/1240, 74.52%). Women constituted 86.61% (1074/1240) of the sample. Participants' ages ranged between 20 and 66 years (mean 36.21, SD 10.18). Nearly 17% of the sample had received specialist support (psychotherapy or pharmacotherapy) to cope with job stress at some point in their professional career.

Preliminary Results

Dropout Analysis and Randomization Check

Out of the 1240 participants, 217 completed the posttest assessment and 147 completed the 6-month follow-up. Intervention dropout, defined as a loss to posttest, was 82.50% (1023/1240). The Little [75] test revealed that data were not missing completely at random (χ^2_{172}=254.9 P<.001). The subsequent ANOVA and chi-square tests confirmed the missing at random (MAR) pattern, depending on the level of primary outcomes at the baseline, condition assignment, and selected job and demographic characteristics. We found that completers were assigned to shorter intervention modules (χ^2_1=14.9; P<.001), were older ($F_{1,1238}$=52.10; P<.001; η^2=0.04), with longer total job tenure ($F_{1, 1238}$=43.58; P<.001; η^2=0.03) and tenure in the current position ($F_{1,1238}$=28.20; P<.001; η^2=0.02), and had been in their respective relationships for longer ($F_{1,1238}$=17.11; P<.001; η^2=0.01). They also showed a higher initial expectancy ($F_{1,1238}$=12.33; P<.001; η^2=0.01) and perceived credibility of the intervention ($F_{1,1238}$=10.57; P=.001; η^2=0.01). Those who completed the intervention presented lower baseline job stress ($F_{1,1238}$=9.05; P=.003; η^2=0.01), job burnout ($F_{1,1238}$=5.44; P=.02; η^2=0.01), depression ($F_{1,1238}$=18.09; P<.001; η^2=0.01), and job-related posttraumatic stress ($F_{1,1238}$=7.91; P=.005; η^2=0.01). No other differences were detected. The MAR pattern of missingness justified conducting MI [74].

Participants randomized into 4 study conditions did not differ with regard to either the study variables at baseline or on descriptive characteristics such as age, tenure, and sex. Finally, participants across the 4 conditions did not differ with regard to their expectations of the treatment or its perceived credibility.

Intervention usage was similar between the experimental and control conditions. In the experimental condition reflecting the cultivation process (self-efficacy precedes perceived social support), 209 participants (out of 311 randomized) engaged in at least one exercise in the first module, and 33 of them engaged in at least one exercise in the second module. For the experimental condition reflecting the enabling process (n=311), these numbers were 206 and 43, respectively. In each control condition (n=309), 205 participants engaged in at least one exercise in the self-efficacy or social support enhancement module. Usage of optional modules was similar across the conditions and varied between 98 participants benefiting from at least one optional module in the social support control condition to 124 participants in the self-efficacy control condition.

Descriptive Statistics

Means, standard deviations, and Pearson correlations for all measures are found in Multimedia Appendix 1.

Effects on Self-Efficacy and Perceived Social Support

We found that there was a significant difference in job stress and burnout-related self-efficacy between time 1 and time 2 (B=−0.08, SE=0.01; 95% CI −0.11 to −0.06) but not between time 2 and time 3 (B=0.00, SE=0.01; 95% CI −0.02 to 0.03). Pairwise comparisons showed that self-efficacy plateaued after posttest but remained significantly higher than at baseline. We observed exactly the same pattern for social support self-efficacy, with a significant difference between time 1 and time 2 (B=−0.18, SE=0.03; 95% CI −0.23 to −0.13) and a nonsignificant difference between time 2 and time 3 (B=0.00, SE=0.02; 95% CI −0.03 to 0.04). In the case of social support, there were significant differences both in the time 1 and time 2 lag (B=−0.19, SE=0.01; 95% CI −0.21 to −0.17) and time 2 and time 3 lag (B=−0.06, SE=0.01; 95% CI −0.08 to −0.03). Pairwise comparisons showed that after social support increased between baseline and posttest, it significantly dropped at the follow-up; however, it remained higher than at baseline.

Hypotheses Testing

First, for each outcome, we compared a reference model with fixed effects of condition assignment, time, and time by condition interaction as well as 2 random effects associated with each participant—intercept and time—with nested models in which either random effect associated with time or random intercept was not included. For all the outcomes, the differences in –2 Restricted Maximum Likelihood log-likelihood between the reference and nested models were statistically significant, and therefore, we retained both random effects. Moreover, for each of the outcomes, we found them to be statistically significant, indicating that within each study condition, there were considerable variations between participants before they started the intervention and in how they responded over time (Table 2).

Table 2. Results of linear mixed effects models.

Outcome	Fixed effects		Random effects	
	Estimate (SE)	95% CI	Estimate (SE)	95% CI
Job stress				
Intercept[a]	2.17-2.23 (0.03)	2.11 to 2.17, 2.22 to 2.45	0.21 (0.01)	0.19-0.23
SE[b] to SE+SS[c]	0.02 (0.04)	−0.06 to 0.10	—[d]	—
SE to SS+SE[e]	−0.03 (0.04)	−0.1 to 0.05	—	—
SS[f] to SE+SS	−0.02 (0.04)	−0.10 to 0.06	—	—
SS to SS+SE	−0.07 (0.04)	−0.15 to 0.01	—	—
SE+SS to SS+SE	0.05 (0.04)	−0.03 to 0.12	—	—
Time[a]	−0.08 to −0.09 (0.01)	−0.11 to −0.10, −0.07 to −0.06	0.01 (0.00)	0.01-0.01
Time×SE to SE+SS	−0.05 (0.01)	−0.08 to −0.03	—	—
Time×SE to SS+SE	0.01 (0.01)	−0.01 to 0.04	—	—
Time×SS to SE+SS	−0.07 (0.01)	−0.09 to−0.04	—	—
Time×SS to SS+SE	0.00 (0.01)	−0.03 to 0.03	—	—
Time×SE+SS to SS+SE	−0.07 (0.01)	−0.09 to −0.04	—	—
Job burnout				
Intercept[a]	2.61-2.70 (0.03)	2.56 to 2.65, 2.67 to 2.76	0.21 (0.01)	0.19-0.24
SE to SE+SS	0.03 (0.04)	−0.05 to 0.11	—	—
SE to SS+SE	0.04 (0.04)	−0.03 to 0.12	—	—
SS to SE+SS	−0.06 (0.04)	−0.14 to 0.02	—	—
SS to SS+SE	−0.05 (0.04)	−0.12 to 0.03	—	—
SE+SS to SS+SE	−0.01 (0.04)	−0.09 to 0.06	—	—
Time[a]	−0.02 to −0.06 (0.01)	−0.08 to −0.03, −0.05 to −0.00	0.01 (0.00)	0.00-0.01
Time×SE to SE+SS	−0.03 (0.01)	−0.05 to −0.00	—	—
Time×SE to SS+SE	−0.04 (0.01)	−0.04 to −0.00	—	—
Time×SS to SE+SS	0.02 (0.01)	−0.00 to 0.05	—	—
Time×SS to SS+SE	0.01 (0.01)	−0.02 to 0.03	—	—
Time×SE+SS to SS+SE	0.02 (0.01)	−0.01 to 0.04	—	—
Work engagement				
Intercept[a]	4.06-4.21 (0.06)	3.95 to 4.09, 4.18 to 4.32	0.89 (0.04)	0.81-0.98
SE to SE+SS	−0.03 (0.08)	−0.19 to 0.13	—	—
SE to SS+SE	−0.04 (0.08)	−0.20 to 0.12	—	—
SS to SE+SS	0.12 (0.08)	−0.04 to 0.28	—	—
SS to SS+SE	0.10 (0.08)	−0.06 to 0.26	—	—
SE+SS to SS+SE	0.01 (0.08)	0.15 to 0.17	—	—
Time[a]	−0.09 to −0.19 (0.02)	−0.23 to −0.13, −0.16 to −0.05	0.03 (0.01)	0.02-0.04
Time×SE to SE+SS	−0.05 (0.03)	−0.10 to 0.01	—	—
Time×SE to SS+SE	−0.02 (0.03)	−0.07 to 0.04	—	—
Time×SS to SE+SS	−0.14 (0.03)	−0.19 to −0.08	—	—
Time×SS to SS+SE	−0.10 (0.03)	−0.16 to −0.05	—	—
Time×SE+SS to SS+SE	−0.03 (0.03)	−0.09 to 0.02	—	—

Outcome	Fixed effects		Random effects	
	Estimate (SE)	95% CI	Estimate (SE)	95% CI
Depression				
Intercept[a]	1.16-1.21 (0.04)	1.08 to 1.13, 1.23 to 1.28	0.36 (0.02)	0.33, 0.39
SE to SE+SS	0.05 (0.05)	−0.06 to 0.15	—	—
SE to SS+SE	−0.05 (0.05)	−0.15 to 0.05	—	—
SS to SE+SS	0.05 (0.05)	−0.05 to 0.15	—	—
SS to SS+SE	−0.05 (0.05)	−0.15 to 0.06	—	—
SE+SS to SS+SE	0.09 (0.05)	−0.01 to 0.20	—	—
Time	−0.04 to −0.07 (0.01)	−0.10 to −0.06, −0.05 to −0.01	0.02 (0.00)	0.02, 0.02
Time×SE to SE+SS	−0.01 (0.01)	−0.05 to 0.02	—	—
Time×SE to SS+SE	0.01 (0.02)	−0.02 to 0.05	—	—
Time×SS to SE+SS	−0.05 (0.02)	−0.08 to −0.01	—	—
Time×SS to SS+SE	−0.02 (0.02)	−0.06 to 0.01	—	—
Time×SE+SS to SS+SE	−0.03 (0.02)	−0.06 to 0.01	—	—
Job-related traumatic stress[g]				
Intercept[a]	1.26-1.38 (0.05)	1.17 to 1.29, 1.36 to 1.48	0.60 (0.03)	0.54, 0.65
SE to SE+SS	−0.10 (0.07)	−0.23 to 0.03	—	—
SE to SS+SE	−0.09 (0.07)	−0.22 to 0.04	—	—
SS to SE+SS	0.02 (0.07)	−0.11 to 0.16	—	—
SS to SS+SE	0.03 (0.07)	−0.10 to 0.16	—	—
SE+SS to SS+SE	−0.01 (0.07)	−0.14 to 0.12	—	—
Time[a]	0.00-0.03 (0.02)	−0.06 to −0.00, 0.01 to 0.07	0.05 (0.00)	0.04, 0.06
Time×SE to SE+SS	−0.02 (0.03)	−0.07 to 0.03	—	—
Time×SE to SS+SE	0.03 (0.03)	−0.02 to 0.08	—	—
Time×SS to SE+SS	−0.07 (0.03)	−0.12 to −0.02	—	—
Time×SS to SS+SE	−0.03 (0.03)	−0.08 to 0.02	—	—
Time×SE+SS to SS+SE	−0.04 (0.03)	−0.09 to 0.01	—	—

[a] As the values differ slightly depending on which study condition is used as a reference category, a range of estimates and confidence intervals are provided.

[b] SE: self-efficacy enhancement module.

[c] SE+SS: self-efficacy and perceived social support sequential enhancement module.

[d] The lack of data in these cells follows statistical analysis: these values are only provided for intercept and time.

[e] SS+SE: perceived social support and self-efficacy sequential enhancement module.

[f] SS: perceived social support enhancement module.

[g] Controlling for exposure to traumatic events.

We expected that the experimental conditions (ie, the ones comprised 2 modules: SE+SS and SS+SE) would be more effective than each active control condition (ie, SE and SS) in reducing job stress and job burnout (primary outcomes) as well as depression and job-related posttraumatic stress (secondary outcomes) at posttest (time 2), and we expected these effects to remain at a 6-month follow-up (time 3). In the case of work engagement, we expected it to be higher at time 2 in the experimental conditions in comparison with active controls, and we expected the effects to remain at a 6-month follow-up (time 3). We were also interested in whether the 2 experimental conditions would differ in their efficacy.

Job Stress

The mixed effects model for job stress showed no significant effect of condition assignment on job stress for any of the between-group comparisons, but it did show a significant effect of time for all comparisons (Table 2), indicating that stress decreased with time. Significant interaction effects were found for time and the comparison between the experimental condition SE+SS with active control SE (B=−0.05, SE=0.01; (95% CI

−0.08 to −0.03) and with active SS control (B=−0.06, SE=0.01; 95% CI −0.09 to −0.04) as well as for time and the comparison between the 2 experimental groups (B=−0.07, SE=0.01; 95% CI −0.09 to −0.04). Bonferroni post hoc tests revealed that at time 2, job stress was significantly lower in the experimental SE+SS condition when compared with the active control SS (*d*=−0.21; 95% CI −0.37 to −0.05). However, at time 3, it was lower in the SE+SS condition than in the remaining 3 conditions: for SS, *d*=−0.41, 95% CI −0.57 to −0.25; for SE, *d*=−0.24, 95% CI −0.40 to −0.09; and for SS+SE, *d*=−0.24, 95% CI −0.39 to −0.08.

Job Burnout

Assignment had no main effect on job burnout for either of the comparisons between study conditions, but there was a main effect of time for all the comparisons (Table 2), indicating that burnout decreased with time. Interaction effects of time and the comparisons between groups were either not significant or marginally significant at a level that did not warrant further investigation due to the risk of inflated type I error resulting from multiple comparisons.

Work Engagement

We found no main effect of condition assignment on work engagement, but we did find a main effect of time (Table 2) for the comparisons between all study conditions, showing that work engagement decreased with time. Significant interactions were found for time and the comparisons between SS active control and 2 experimental groups: SE+SS (B=−0.13, SE=0.03; 95% CI −0.19 to −0.08) and SS+SE (B=−0.10, SE=0.03; 95% CI −0.16 to −0.05). Bonferroni post hoc tests demonstrated that there were no significant differences between study conditions at time 2, but at time 3, work engagement was significantly lower in the experimental condition SE+SS than in the active control SS (*d*=−0.20; 95% CI −0.36 to −0.04).

Depression

Similar to the case of work engagement, for depression, we found an effect of time, indicating that depression decreased with time, but no effect of condition assignment for any of the between-group comparisons (Table 2). The only significant interaction was found for time and the comparison between active control SS and experimental condition SE+SS (B=−0.05, SE=0.02; 95% CI −0.08 to −0.01). However, Bonferroni post hoc tests revealed that there were no meaningful differences between these conditions at either posttest or at follow-up.

Job-Related Traumatic Stress

For job-related traumatic stress, we additionally controlled for the fixed effect of the exposure to traumatic events. Neither condition assignment nor time had an effect on the outcome (Table 2). A significant interaction was observed for time and the comparison between active control SS and experimental SE+SS condition (B=−0.07, SE=0.03; 95% CI −0.12 to −0.02). Bonferroni post hoc analysis showed that this difference was significant only at time 3, with lower stress in the experimental condition (*d*=−0.24; 95% CI −0.40 to −0.08).

Predictors of Outcomes

To identify potential predictors of the 2 primary outcomes, job stress and job burnout, we ran bootstrapped regression analyses with the outcomes reflecting change between pretest and posttest (residual change score) and 95% bias-corrected bootstrap CIs. Age, gender, marital status, average weekly work hours, type of organization (public vs private), receiving treatment in the past, and general tenure were not associated with changes in job stress and job burnout. However, tenure in the current workplace turned out to be a significant predictor of both change in job stress (β=−.22; 95% CI −0.02 to −0.01) and in burnout (β=−.18; 95% CI −0.02 to −0.01): greater reduction in both outcomes was observed among participants with shorter tenure. Changes in job stress and burnout were also associated with the level of expectations of the intervention. Participants who had higher expectations of the treatment and perceived the intervention as more credible experienced a greater decrease in stress (credibility: β=.16; 95% CI 0.04-0.09 and expectancy: β=.17; 95% CI 0.04-0.09) and in burnout (credibility: β=.10; 95% CI 0.01-0.06 and expectancy: β=.25; 95% CI 0.06-0.10).

In terms of using the intervention, change in job stress was predicted by engaging in exercises in the self-efficacy enhancing module (β=.11; 95% CI 0.06-0.22): at the posttest, stress was lower among the participants who engaged in the self-efficacy module. Using social support enhancing exercises and optional modules (relaxation, mindfulness, cognitive restructuring, and lifestyle) was not associated with change in job stress. A change in job burnout was predicted neither by engagement in self-efficacy nor by social support exercises. Only using optional modules was associated with a reduction in job burnout at the posttest (β=.07; 95% CI 0.01-0.16).

Discussion

Principal Findings

This study had 2 objectives. First, we wanted to test whether enhancing 2 psychological resources, perceived social support and self-efficacy, would be more effective for improving the well-being of medical professionals than strengthening only 1 of them. Second, we aimed to experimentally verify the enabling versus the cultivation hypothesis, that is, to test whether the sequence in which social support and self-efficacy were targeted would have an effect on well-being. To reflect the complexity of a natural occupational context, we tested the efficacy of the intervention for 5 outcomes. We found that job burnout did not depend on assignment to the study condition. However, allocation to study condition did have an effect on another primary outcome, job stress. Immediately after the intervention, we found that one experimental condition (ie, the one reflecting the cultivation hypothesis by first enhancing self-efficacy and later social support) was more effective in reducing stress than the control condition that solely targeted social support, but not the one dedicated to enhancing only self-efficacy. This result partially supported our expectations. However, 6 months later, the same experimental condition was found to be the most effective. This result supported the cultivation process of stress reduction over an enabling one. When we analyzed how participants used the Med-Stress intervention regardless of condition assignment, we found that there was a reduction in job burnout, and it was predicted by using optional modules that comprised CBT-based exercises. Perhaps these exercises, and not those dedicated to building personal resources (ie, self-efficacy and perceived social support), were responsible for the decrease in burnout, albeit small, in participants across all conditions. Job stress, on the other hand, was not associated with completing optional modules; the only significant predictor of its decrease was using self-efficacy exercises.

No meaningful differences between study conditions were detected for depression. In the case of job-related traumatic stress, between-group differences at the posttest were negligible, whereas at the follow-up, we again found that participants in the condition reflecting the cultivation hypothesis reported greater reduction in stress than those in a control condition enhancing solely social support, although the effect was small. Work engagement also revealed different patterns at posttest and follow-up. Immediately after the intervention, there were no meaningful differences in work engagement between study conditions; however, 6 months later, work engagement was lower in the same experimental condition reflecting the cultivation hypothesis than in the control condition that aimed to build only social support. Moreover, contrary to what might be expected, work engagement decreased over time. In fact, at the follow-up, work engagement was significantly lower in the same condition that we found was beneficial for the reduction of both types of job-related stress. In other words, participants who first completed self-efficacy enhancing exercises and then the ones dedicated to building perception of social support reported lower stress and lower work engagement. This pattern of results suggests that a decrease in work engagement might be beneficial to medical professionals. In fact, although work engagement is overly considered a positive state, associated with numerous beneficial consequences for a person, such as better self-reported health, a so-called dark side of engagement has also been identified, suggesting that over-engaged workers might experience undesired outcomes [76]. Moreover, the results of a recent meta-analysis [77] showed that only half of the interventions aimed at improving work engagement were found to have a positive effect. The rest had no effect or had a negative effect, although the latter were in the minority. Interestingly, the intervention that resulted in decreased work engagement was dedicated to service workers in older care, a group that is similar to the one in Med-Stress [78]. This might indicate that, at least for these occupational groups, being highly engaged in work can be detrimental. At the same time, this interpretation needs to be treated with caution as one of the options that require further analysis.

Taken together, these results show that the experimental condition reflecting the cultivation hypothesis—self-efficacy preceding social support—is the most beneficial in the long run, although this finding needs to be treated with caution for the following reasons: first, it was not found for all components of well-being. Second, for those that it was found (job stress and traumatic stress), the effect sizes were small. Third, in the case of work engagement, more research is needed to establish whether the decrease over time was beneficial or detrimental to the participants. Small effect sizes are perhaps not that surprising considering that gain cycles, contrary to loss spirals, are considered to be weaker and take longer to unfold [23]. As 6 months is a relatively short period, subsequent measurement points are needed, and therefore, we have scheduled another follow-up in the future. In addition, small effect sizes could be attributed to the comparisons being made between experimental conditions and active, not passive, controls [79].

Perhaps reaching a result that is not completely coherent for all outcomes is not that surprising when these outcomes are numerous and represent different aspects of well-being. By targeting multiple outcomes with one program, we risk losing precision, but we gain a more naturalistic intervention.

Strengths and Limitations

Med-Stress responds to a call “to conduct feasibility studies and randomized controlled studies on the effect of low-intensity interventions and technology supported (eg, web-based) interventions in low- and middle-income countries (LMICs), preferably using an active control condition as comparison, to ensure we disseminate effective treatments in LMICs” [80]. LMICs are reported to experience a disparity between the need for mental health services and their provision. We would argue that the need for such treatments is also urgent among certain groups, including health care providers. Content development of Med-Stress was preceded by a web-based focus study to better understand the working conditions of the target audience. We found that despite high levels of stress and burnout among medical professionals, no support programs were offered at the organizations where responders worked. Benefiting from traditional help often proves difficult for them due to irregular working schedules and an insufficient number of professionals. Thus, the availability of an internet intervention of low intensity and focused on personal resources that are context-free could be beneficial in alleviating some work-related mental health issues. This became particularly salient a few months later, when the COVID-19 pandemic broke and medical professionals found themselves in need of support under conditions where it could not be easily provided. In fact, to respond to this crisis, we developed a shorter, contextualized version of the Med-Stress intervention, called Med-Stress SOS [81].

Our study has several limitations. First, the dropout was high, which affects the interpretation of the results. We mitigated this risk by applying the MI method. There is an ongoing argument regarding the threshold over which the dropout rate is deemed unacceptable; however, the latest simulations indicate that it is not the proportion but the pattern of missingness that should be taken into account when considering data imputation. Specifically, in the case of the MAR pattern, the MI procedure, which includes all identified variables that differentiate dropouts from completers as imputation predictors, leads to the least biased results [74,82]. As the pattern of missingness in this study was indeed MAR, we followed these recommendations when conducting MI. Moreover, it should be said that internet interventions, in particular self-guided ones, do tend to have high dropout rates. Although such a loss to follow-up is rare, it is not unprecedented [83,84]. However, despite our attempts to mitigate the risks resulting from high dropout, the obtained results need to be treated with caution. This loss reflects the pragmatic nature of this trial: it was conducted in Poland, where internet interventions are still rare, and we suspect that users did not have a framework into which this form of psychological help could be easily incorporated. A high recruitment rate compared with high dropout probably reflects participants' initial enthusiasm and curiosity that diminished over time. This was also reflected in the poor usage of the intervention, particularly in the experimental conditions in which the second assigned module was used much less frequently than the first one. In fact, comments that we received post intervention seem to support this notion: users referred to the content of the exercises (not user-friendly enough), pace of the intervention (new exercise released each week), and the fact that it was a web-based intervention and not an app-based intervention. High dropout was also a reason why we could not conduct a per protocol analysis: the sample of completers was too small to provide sufficient power. Ultimately, we had enough power to detect effects of minimum d=0.21, and therefore, all findings could not be generalized; however, the effects of job stress and job-related traumatic stress were large enough. Some of the phenomena in this study, in particular job burnout and work engagement, are multidimensional, and analyzing them as such could provide a more in-depth understanding of the intervention's impact on them. However, this would further increase the already high number of comparisons that needed to be accounted for in the study design.

Conclusions

This study offers both theoretical and practical contributions. It is a pragmatic trial that offers an insight into how people really use this intervention [85]. It is an experimental verification of the cultivation versus enabling process of stress reduction, demonstrating cautious support for the former. Moreover, we empirically tested the concept of gain cycles for 5 components of well-being. In practical terms, Med-Stress is a support program that was found to be effective in enhancing the well-being of medical professionals. This intervention has the advantage of being broadly accessible to health care workers who currently receive no help in coping with their mental health problems.

Acknowledgments

The authors are grateful to Malwina Puchalska-Kaminska and Krzysztof Rzenca for their assistance in conducting the preimplementation study. The authors thank Malgorzata Zalewska for designing the optional modules and Dr Agnieszka Popiel and Dr Ewa Praglowska for their critical review of the entire content of the intervention. Digitalization of Med-Stress was possible thanks to ongoing support from George Vlaescu. This study was funded by The National Centre for Research and Development in Poland and coordinated by the Central Institute for Labour Protection, National Research Institute in Poland (project no. I.N.16).

Authors' Contributions

ES, ML and RC conceptualized the study. ES and ML conducted statistical analyses and prepared original draft. GA, PC, and RC reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Means, SDs, and correlations for study variables.
[DOCX File , 25 KB - jmir_v23i1e21445_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1585 KB - jmir_v23i1e21445_app2.pdf]

References

1. Wright TA, Cropanzano R. Psychological well-being and job satisfaction as predictors of job performance. J Occup Health Psychol 2000 Jan;5(1):84-94. [doi: 10.1037//1076-8998.5.1.84] [Medline: 10658888]
2. Darr W, Johns G. Work strain, health, and absenteeism: a meta-analysis. J Occup Health Psychol 2008 Oct;13(4):293-318. [doi: 10.1037/a0012639] [Medline: 18837626]
3. Fried Y, Shirom A, Gilboa S, Cooper CL. The mediating effects of job satisfaction and propensity to leave on role stress-job performance relationships: Combining meta-analysis and structural equation modeling. Int J Stress Manag 2008 Nov;15(4):305-328. [doi: 10.1037/a0013932]
4. Brand SL, Coon JT, Fleming LE, Carroll L, Bethel A, Wyatt K. Whole-system approaches to improving the health and wellbeing of healthcare workers: a systematic review. PLoS One 2017;12(12):e0188418 [FREE Full text] [doi: 10.1371/journal.pone.0188418] [Medline: 29200422]
5. Rothenberger DA. Physician burnout and well-being: a systematic review and framework for action. Dis Colon Rectum 2017 Jun;60(6):567-576. [doi: 10.1097/DCR.0000000000000844] [Medline: 28481850]

6. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. Lancet 2016 Nov 5;388(10057):2272-2281. [doi: 10.1016/S0140-6736(16)31279-X] [Medline: 27692469]
7. Rickard G, Lenthall S, Dollard M, Opie T, Knight S, Dunn S, et al. Organisational intervention to reduce occupational stress and turnover in hospital nurses in the Northern territory, Australia. Collegian 2012;19(4):211-221. [doi: 10.1016/j.colegn.2012.07.001] [Medline: 23362607]
8. Nielsen K, Nielsen MB, Ogbonnaya C, Känsälä M, Saari E, Isaksson K. Workplace resources to improve both employee well-being and performance: a systematic review and meta-analysis. Work & Stress 2017 Mar 24;31(2):101-120. [doi: 10.1080/02678373.2017.1304463]
9. Joyce S, Modini M, Christensen H, Mykletun A, Bryant R, Mitchell PB, et al. Workplace interventions for common mental disorders: a systematic meta-review. Psychol Med 2016 Mar 06;46(4):683-697. [doi: 10.1017/S0033291715002408] [Medline: 26620157]
10. Lomas T, Medina JC, Ivtzan I, Rupprecht S, Eiroa-Orosa FJ. Mindfulness-based interventions in the workplace: an inclusive systematic review and meta-analysis of their impact upon wellbeing. J Posit Psychol 2018 Sep 30;14(5):625-640. [doi: 10.1080/17439760.2018.1519588]
11. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
12. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
13. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
14. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
15. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
16. Geraedts AS, Kleiboer AM, Twisk J, Wiezer NM, van Mechelen W, Cuijpers P. Long-term results of a web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014 Jul 9;16(7):e168 [FREE Full text] [doi: 10.2196/jmir.3539] [Medline: 25008127]
17. Aikens KA, Astin J, Pelletier KR, Levanovich K, Baase CM, Park YY, et al. Mindfulness goes to work: impact of an online workplace intervention. J Occup Environ Med 2014 Jul;56(7):721-731. [doi: 10.1097/JOM.0000000000000209] [Medline: 24988100]
18. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
19. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
20. Lazarus R, Folkman S. Stress, Appraisal, and Coping. New York: Springer Pub. Co; 1984.
21. Demerouti E, Nachreiner F, Bakker AB, Schaufeli WB. The Job Demands-Resources Model of Burnout. Journal of Applied Psychology 2001 Jun;86(3):499-512. [doi: 10.1037/0021-9010.86.3.499] [Medline: 11419809]
22. Hobfoll SE. The influence of culture, community, and the nested‐self in the stress process: advancing conservation of resources theory. Appl Psychol 2001 Dec 21;50(3):337-421. [doi: 10.1111/1464-0597.00062]
23. Hobfoll SE, Halbesleben J, Neveu J, Westman M. Conservation of resources in the organizational context: the reality of resources and their consequences. Annu Rev Organ Psychol Organ Behav 2018 Jan 21;5(1):103-128. [doi: 10.1146/annurev-orgpsych-032117-104640]
24. Shoji K, Lesnierowska M, Smoktunowicz E, Bock J, Luszczynska A, Benight CC, et al. What comes first, job burnout or secondary traumatic stress? Findings from two longitudinal studies from the U.S. And Poland. PLoS One 2015;10(8):e0136730 [FREE Full text] [doi: 10.1371/journal.pone.0136730] [Medline: 26305222]
25. Llorens S, Schaufeli W, Bakker A, Salanova M. Does a positive gain spiral of resources, efficacy beliefs and engagement exist? Comput Hum Behav 2007 Jan;23(1):825-841. [doi: 10.1016/j.chb.2004.11.012]
26. Salanova M, Schaufeli WB, Xanthopoulou D, Bakker AB. The gain spiral of resources and work engagement: Sustaining a positive worklife. In: Bakker AB, Leiter MP, editors. Work Engagement: A Handbook of Essential Theory and Research. Milton Park, Oxfordshire: Psychology Press; 2010.

27. Seppälä P, Hakanen J, Mauno S, Perhoniemi R, Tolvanen A, Schaufeli W. Stability and change model of job resources and work engagement: a seven-year three-wave follow-up study. Eur J Work Organ Psychol 2014 May 7;24(3):360-375. [doi: 10.1080/1359432X.2014.910510]
28. Kim SD, Hollensbe EC, Schwoerer CE, Halbesleben JRB. Dynamics of a wellness program: a conservation of resources perspective. J Occup Health Psychol 2015 Jan;20(1):62-71. [doi: 10.1037/a0037675] [Medline: 25151464]
29. Nurullah A. Received and provided social support: a review of current evidence and future directions. Am J Health Stud 2012;27(3):173-188 [FREE Full text]
30. Sarason BR, Sarason IG, Pierce GR, editors. Traditional views of social support and their impact on assessment. In: Social Support: An Interactional View. Wiley series on personality processes. Hoboken, New Jersey: John Wiley & Sons; 1990.
31. Haber MG, Cohen JL, Lucas T, Baltes BB. The relationship between self-reported received and perceived social support: a meta-analytic review. Am J Community Psychol 2007 Mar;39(1-2):133-144. [doi: 10.1007/s10464-007-9100-9] [Medline: 17308966]
32. Hobfoll SE, Freedy J, Lane C, Geller P. Conservation of social resources: social support resource theory. J Soc Pers Relatsh 2016 Jun 29;7(4):465-478. [doi: 10.1177/0265407590074004]
33. Bandura A. Self-Efficacy: The Exercise of Control. New York City: Worth Publishers; 1997.
34. Byars-Winston A, Diestelmann J, Savoy JN, Hoyt WT. Unique effects and moderators of effects of sources on self-efficacy: a model-based meta-analysis. J Couns Psychol 2017 Nov;64(6):645-658. [doi: 10.1037/cou0000219] [Medline: 29154576]
35. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. New Jersey: Prentice-Hall, Inc; 1986.
36. Schwarzer R, Knoll N. Functional roles of social support within the stress and coping process: a theoretical and empirical overview. Int J Psychol 2007 Aug;42(4):243-252. [doi: 10.1080/00207590701396641]
37. Hohl DH, Schultze M, Keller J, Heuse S, Luszczynska A, Knoll N. Inter-relations between partner-provided support and self-efficacy: a dyadic longitudinal analysis. Appl Psychol Health Well Being 2019 Nov;11(3):522-542. [doi: 10.1111/aphw.12166] [Medline: 31231970]
38. Benight CC, Bandura A. Social cognitive theory of posttraumatic recovery: the role of perceived self-efficacy. Behav Res Ther 2004 Oct;42(10):1129-1148. [doi: 10.1016/j.brat.2003.08.008] [Medline: 15350854]
39. Shoji K, Bock J, Cieslak R, Zukowska K, Luszczynska A, Benight CC. Cultivating secondary traumatic growth among healthcare workers: the role of social support and self-efficacy. J Clin Psychol 2014 Sep;70(9):831-846. [doi: 10.1002/jclp.22070] [Medline: 24420964]
40. Parent-Thirion A, Biletta I, Cabrita J, Aleksandra W, Oscar V. 6th European Working Conditions Survey: Overview Report 2017 update. Luxembourg: Publications Office of the European Union; 2017.
41. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. Lancet 2009 Nov 14;374(9702):1714-1721. [doi: 10.1016/S0140-6736(09)61424-0] [Medline: 19914516]
42. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
43. International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision). World Health Organization. 2018. URL: https://icd.who.int/browse11/l-m/en
44. Maslach C, Schaufeli WB, Leiter MP. Job burnout. Annu Rev Psychol 2001;52:397-422. [doi: 10.1146/annurev.psych.52.1.397] [Medline: 11148311]
45. Schaufeli WB, Leiter MP, Maslach C. Burnout: 35 years of research and practice. Career Dev Int 2009 Jun 19;14(3):204-220. [doi: 10.1108/13620430910966406]
46. Zgliczyńska M, Zgliczyński S, Ciebiera M, Kosińska-Kaczyńska K. Occupational burnout syndrome in Polish physicians: a systematic review. Int J Environ Res Public Health 2019 Dec 10;16(24):- [FREE Full text] [doi: 10.3390/ijerph16245026] [Medline: 31835554]
47. Dubale BW, Friedman LE, Chemali Z, Denninger JW, Mehta DH, Alem A, et al. Systematic review of burnout among healthcare providers in sub-Saharan Africa. BMC Public Health 2019 Sep 11;19(1):1247 [FREE Full text] [doi: 10.1186/s12889-019-7566-7] [Medline: 31510975]
48. Imai H, Nakao H, Nakagi Y, Niwata S, Sugioka Y, Itoh T, et al. Prevalence of burnout among public health nurses in charge of mental health services and emergency care systems in Japan. Environ Health Prev Med 2006 Nov;11(6):286-291 [FREE Full text] [doi: 10.1007/BF02898018] [Medline: 21432357]
49. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. J Am Med Assoc 2018 Sep 18;320(11):1131-1150 [FREE Full text] [doi: 10.1001/jama.2018.12777] [Medline: 30326495]
50. Schaufeli W, Salanova M, González-romá V, Bakker A. The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. J Happiness Stud 2002;3(1):71-92.
51. Moeller J, Ivcevic Z, White A, Menges J, Brackett M. Highly engaged but burned out: intra-individual profiles in the US workforce. Career Dev Int 2018 Feb 6;23(1):86-105. [doi: 10.1108/CDI-12-2016-0215]
52. American Psychiatric Association. Diagnostic and statistical manual of mental disorders DSM-5. Washington, DC: American Psychiatric Association; 2013.

53. Beck CT. Secondary traumatic stress in nurses: a systematic review. Arch Psychiatr Nurs 2011 Feb;25(1):1-10. [doi: 10.1016/j.apnu.2010.05.005] [Medline: 21251596]
54. Nimmo A, Huggard P. A systematic review of the measurement of compassion fatigue, vicarious trauma, and secondary traumatic stress in physicians. Australas J Disaster Trauma Stud 2013;1:37-44.
55. Harvey SB, Glozier N, Henderson M, Allaway S, Litchfield P, Holland-Elliott K, et al. Depression and work performance: an ecological study using web-based screening. Occup Med (Lond) 2011 May;61(3):209-211. [doi: 10.1093/occmed/kqr020] [Medline: 21525074]
56. Beiwinkel T, Eißing T, Telle N, Siegmund-Schultze E, Rössler W. Effectiveness of a web-based intervention in reducing depression and sickness absence: randomized controlled trial. J Med Internet Res 2017 Jun 15;19(6):e213 [FREE Full text] [doi: 10.2196/jmir.6546] [Medline: 28619701]
57. van Dam A. Subgroup analysis in burnout: relations between fatigue, anxiety, and depression. Front Psychol 2016;7:90 [FREE Full text] [doi: 10.3389/fpsyg.2016.00090] [Medline: 26869983]
58. Smoktunowicz E, Lesnierowska M, Cieslak R, Carlbring P, Andersson G. Efficacy of an Internet-based intervention for job stress and burnout among medical professionals: study protocol for a randomized controlled trial. Trials 2019 Jun 10;20(1):338 [FREE Full text] [doi: 10.1186/s13063-019-3401-9] [Medline: 31182128]
59. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
60. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
61. Magnusson, K. Power Analysis for Longitudinal 2- and 3-Level Models Challenges and Some Solutions Using the R Package powerlmm. R Psychologist. 2018. URL: https://rpsychologist.com/powerlmm-intro-slides [accessed 2020-12-15]
62. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interv 2016 Nov;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]
63. Demerouti E, Mostert K, Bakker AB. Burnout and work engagement: a thorough investigation of the independency of both constructs. J Occup Health Psychol 2010 Jul;15(3):209-222. [doi: 10.1037/a0019408] [Medline: 20604629]
64. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
65. Schaufeli WB, Shimazu A, Hakanen J, Salanova M, De Witte H. An ultra-short measure for work engagement. Eur J Psychol Assess 2019 Jul;35(4):577-591. [doi: 10.1027/1015-5759/a000430]
66. Weathers F, Litz B, Keane T, Palmieri P, Marx B, Schnurr P. PTSD Checklist for DSM-5 (PCL-5). PTSD: National Center for PTSD. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2020-12-15]
67. Lua H. The Mediating Role of Work Stress and Burnout Management Self-efficacy in the Job Demand - Resources Model [dissertation]. Warsaw, Poland: SWPS University of Social Sciences and Humanities; 2008.
68. Schwarzer R, Schulz U. Berlin Social Support Scales (BSSS). Measurement Instrument Database for the Social Sciences. 2000. URL: https://www.midss.org/content/berlin-social-support-scales-bsss [accessed 2020-12-15]
69. Widerszal-Bazyl M, Cieślak R. Monitoring psychosocial stress at work: development of the psychosocial working conditions questionnaire. Int J Occup Saf Ergon 2000;Spec No:59-70. [doi: 10.1080/10803548.2000.11105108] [Medline: 10828153]
70. Cieslak R, Shoji K, Luszczynska A, Taylor S, Rogala A, Benight CC. Secondary trauma self-efficacy: concept and its measurement. Psychol Assess 2013 Sep;25(3):917-928. [doi: 10.1037/a0032687] [Medline: 23647049]
71. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
72. West B, Welch K, Galecki A. Linear Mixed Models. A practical guide using statistical software, Second Edition. Boca Raton, Florida: Chapman and Hall/CRC; 2015:9781466560994.
73. Steketee G, Chambless D. Methodological issues in prediction of treatment outcome. Clin Psychol Rev 1992;12(4):387-400 [FREE Full text] [doi: 10.1016/0272-7358(92)90123-P]
74. Groenwold RH, Donders AR, Roes KC, Harrell FE, Moons KG. Dealing with missing outcome data in randomized trials and observational studies. Am J Epidemiol 2012 Feb 1;175(3):210-217. [doi: 10.1093/aje/kwr302] [Medline: 22262640]
75. Little R, Rubin D. Statistical Analysis with Missing Data (2 edition). Hoboken, New Jersey: Wiley; 2002.
76. Bakker AB, Albrecht SL, Leiter MP. Key questions regarding work engagement. Eur J Work Organ Psychol 2011 Feb;20(1):4-28. [doi: 10.1080/1359432x.2010.485352]
77. Knight C, Patterson M, Dawson J. Work engagement interventions can be effective: a systematic review. Eur J Work Organ Psychol 2019 Apr 6;28(3):348-372. [doi: 10.1080/1359432x.2019.1588887]
78. Ng S. Is brief daily body–mind–spirit practice desirable for staff who provide services for elderly people? Two pilot studies with care and professional workers. Asia Pac J Soc Work Dev 2014 Apr 25;24(4):227-237. [doi: 10.1080/02185385.2013.843471]
79. Emanuel EJ, Miller FG. The ethics of placebo-controlled trials--a middle ground. N Engl J Med 2001 Sep 20;345(12):915-919. [doi: 10.1056/NEJM200109203451211] [Medline: 11565527]

80. Bockting CL, Williams AD, Carswell K, Grech AE. The potential of low-intensity and online interventions for depression in low- and middle-income countries. Glob Ment Health (Camb) 2016;3:e25 [FREE Full text] [doi: 10.1017/gmh.2016.21] [Medline: 28596893]
81. Smoktunowicz E, Lesnierowska M, Cieslak R, Carlbring P, Andersson G. Med-Stress SOS: Internet Intervention for Medical Professionals Working During Health Crisis. Center for Open Science. 2020. URL: osf.io/yx4tm [accessed 2020-12-15]
82. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol 2019 Jun;110:63-73 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.02.016] [Medline: 30878639]
83. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
84. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
85. Tosh G, Soares-Weiser K, Adams CE. Pragmatic vs explanatory trials: the pragmascope tool to help measure differences in protocols of mental health randomized controlled trials. Dialogues Clin Neurosci 2011;13(2):209-215 [FREE Full text] [Medline: 21842618]

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
COR: conservation of resources theory
LMIC: low- and middle-income country
MAR: missing at random
MI: multiple imputation
RCT: randomized controlled trial
SE: self-efficacy enhancement module
SE+SS: self-efficacy and perceived social support sequential enhancement module
SS: perceived social support enhancement module
SS+SE: perceived social support and self-efficacy sequential enhancement module

Edited by G Eysenbach; submitted 16.06.20; peer-reviewed by R Rahman, M Lozano-Lozano; comments to author 28.10.20; accepted 28.10.20; published 11.01.21.

Please cite as:
Smoktunowicz E, Lesnierowska M, Carlbring P, Andersson G, Cieslak R
Resource-Based Internet Intervention (Med-Stress) to Improve Well-Being Among Medical Professionals: Randomized Controlled Trial
J Med Internet Res 2021;23(1):e21445
URL: https://www.jmir.org/2021/1/e21445
doi:10.2196/21445
PMID:33427674

Original Paper

Effects of Website Interactivity on Skin Cancer–Related Intentions and User Experience: Factorial Randomized Experiment

Zhaomeng Niu[1], PhD; Jessica Fitts Willoughby[2], PhD; Elliot J Coups[3], PhD; Jerod L Stapleton[4], PhD

[1]Section of Behavioral Sciences, Division of Medical Oncology, Rutgers Institute of New Jersey, New Brunswick, NJ, United States
[2]The Edward R Murrow College of Communication, Washington State University, Pullman, WA, United States
[3]Medical Data Analytics, Parsippany, NJ, United States
[4]Department of Health, Behavior & Society, College of Public Health, University of Kentucky, Lexington, KY, United States

Corresponding Author:
Zhaomeng Niu, PhD
Section of Behavioral Sciences
Division of Medical Oncology
Rutgers Institute of New Jersey
195 Little Albany Street
Room 5549-20
New Brunswick, NJ, 08903
United States
Phone: 1 5093393076
Email: niuzm321@gmail.com

Abstract

Background: Digital media technologies provide users with the ability to interact with content and to receive information based on their preferences and engagement.

Objective: We used skin cancer and sun protection as a health topic to explore how modality interactivity, interface tools that afford users greater activity, resulting in greater depth and breadth of mentally representing and experiencing mediated content, and message interactivity, the extent to which the system allows users to exchange messages back and forth on health websites, influenced users' attitudes, knowledge, behavioral intentions, and experience.

Methods: We employed a 2×2 (modality interactivity: high vs low; message interactivity: high vs low) between-subject online experiment for which 4 websites were created. Participants (n=293) were recruited using Amazon Mechanical Turk and randomly assigned into to 1 of 4 conditions. After browsing the website, participants completed an online survey regarding their experience and cognitive perceptions. General linear models and path analysis were used to analyze the data.

Results: Both modality interactivity ($P=.001$) and message interactivity ($P<.001$) had an impact on intention to use sun protection. Attitudes toward health websites and perceived knowledge mediated the effects of modality interactivity and message interactivity on sun protection use intention, individually. Participants in the high modality interactivity and high message interactivity condition felt more satisfied ($P=.02$). Participants in the low message interactivity condition had more interest in the experience with health websites than participants in the high message interactivity condition ($P=.044$).

Conclusions: Findings suggested that modality interactivity influenced intention to use sun protection directly as well as via attitudes toward the websites. Message interactivity impacted intention to use sunscreen directly and also through perceived knowledge. Implications for designing health websites and health intervention content are discussed.

(J Med Internet Res 2021;23(1):e18299) doi:10.2196/18299

KEYWORDS

behavioral intention; computer-mediated communication; skin cancer; interactivity; user experience

Introduction

Background

The use of technology in communication is ubiquitous in today's society. As a result, communication has shifted from traditional 1-way communication to 2-way reciprocal approaches [1,2]. Media users become active information seekers instead of passive information receivers. The integration of advanced elements of interpersonal communication and mass communication into internet-based communication has resulted in immediate, back-and-forth, customized responses [3,4]. Moreover, users can generate content and share experiences with other users, and users themselves may become sources as well as receivers of information [5,6]. This 2-way flow of communication has been useful for health promotion efforts [7-13].

According to the Pew Internet & American Life Project [14], 1 in 3 Americans have used the internet to solve a medical problem. Due to the public's increasing willingness to be more responsible for their own health [15,16], how to effectively deliver health information using technologies has become a question requiring further investigation. Internet-based interventions were found to work better than noninternet-based interventions for improving a variety of health-related attitudes and behavioral outcomes including increased knowledge, participation in health care, and more [17].

Interactivity, defined as how responsive a system is to a user [18], has been examined by many researchers in health communication as a key characteristic of interactive digital media technologies [19-24]. While some have found that higher levels of interactivity can improve knowledge [20], attitude toward health websites [21,24], and intentions to use a health information resource [25,26], others have found that higher levels of interactivity did not have positive effects on knowledge [22,27] or self-efficacy [22]. Overloading of technological features on media interfaces (exceeding the cognitive processing boundary) may lead to distraction and reactance to the persuasive messages [28-30]. Since there are both arguments for and against using interactive technological affordances, in terms of information processing and assessment, it is important to further examine the role of interactive features in an online environment.

We aimed to examine (1) how different types of interactivity influence health behavioral intentions and whether there is an interaction effect among these 2 types of interactivity on behavioral intentions; (2) how different types of interactivity impact behavioral intentions through different cognition; (3) how different types of interactivity influence individual user experience with health websites.

Skin Cancer and Sun Protection

Skin cancer is the most common cancer in the United States [31]. One in 5 Americans will develop skin cancer in their lifetime [32]. Each year, more than 3.5 million new cases of skin cancer are diagnosed in the United States, which is more than the yearly total of new cases for all other types of cancers [33,34]. However, skin cancer is the most preventable cancer compared to other forms of cancer [35]. The most preventable risk factor is exposure to ultraviolet light [35]. Therefore, most skin cancer interventions aim to promote sun protection behaviors, such as using protective clothes or shades and increasing sunscreen use in the general public or in people at high risk of skin cancer or decreasing exposure to artificial ultraviolet light (eg, tanning beds). Researchers have been working on building effective skin cancer interventions for decades [36-43]; however, no previous study has employed skin cancer as a health context to examine how different types of interactivity affect user experience with health websites, behavioral intentions, and information processing [21].

Interactivity

According to the Modality, Agency, Interactivity, and Navigability (MAIN) model [6], interactivity should be a system of technological affordances that can allow users to make changes to sources, messages, and media while interacting with the interfaces. Based on the above, there are 3 types of interactivity—modality interactivity, message interactivity, and source interactivity—and they can influence user engagement then sequentially affect cognition, attitude, and behavior [6,44].

When interactivity is assessed as a functional modality on the medium interface, it is called *modality interactivity*. Modality interactivity refers to "interface tools that afford users greater activity, resulting in greater depth and breadth of mentally representing and experiencing mediated content [44]." Traditional media usually contain a single modality. For example, print media, may solely have text, and radio may solely provide audio. Digital media have multiple modalities as multimedia content [45,46]. Modality cues not only include video, image, text, and audio but also include new interactive affordances such as hyperlinks, clicking, zooming, dragging, scrolling, and mouse-overs.

Message interactivity has been demonstrated by many researchers through the concept of message exchange [18] or 2-way communication [47]. *Message interactivity* refers to the extent to which the system allows users to exchange messages back and forth. The action of message interactivity is performed through the principle of contingency, which means that "the idea that a given message is contingent upon user reception of the previous message and the ones preceding that [44]." If a system or a media channel allows users to have back and forth interaction in a highly logical flow, the system or the channel is seen as having high message interactivity. Hyperlinks and buttons embedded in websites are a typical format for manipulating message interactivity. Sundar et al [48] examined the effect of how hypertext, when operationalized as message interactivity, allowed users to access information through nonlinear communication while exploring a website.

Previous studies [49-52] have found that modality interactivity can influence cognitive perceptions including attitudes and behavioral intentions. Many empirical studies [20,24,53] have shown the effectiveness of modality interactivity in health communication, such as increasing attitudes toward health websites and physical activity intentions. However, few studies have specifically investigated message interactivity on attitudes and intentions under the context of health. One study [44] found that higher message interactivity can lead to higher evaluation of the content and result in higher intentions to recommend the website to others.

Research Questions and Hypotheses

Based on previously reviewed literature, prior research mainly focused on the effectiveness of interactivity and compare the effects of different levels of interactivity; little research has been conducted to examine the interaction effects of different types of interactivity and whether there is an interaction effect between different types of interactivity. However, in real-life situations, different types of interactivity usually are presented together on the interface, and individuals interact with multiple technological affordances back and forth, curvilinearly. Furthermore, there has been limited research investigating how different types of interactivity influence user experience factors, such as satisfaction and interest.

Thus, we used skin cancer as the context and proposed the following hypotheses and research questions—

Hypothesis 1: Modality interactivity will be positively associated with attitudes toward a health website and intention to use sun protection.

Hypothesis 2: Message interactivity will be positively associated with knowledge of skin cancer and intention to use sun protection.

Research question 1: Is there an interaction effect on attitudes, knowledge or intention?

Hypothesis 3: Modality interactivity will mediate behavioral intention via its prior effects on attitudes toward health websites.

Hypothesis 4: Message interactivity will mediate behavioral intention via its prior effects on knowledge.

Research question 2: How will modality interactivity and message interactivity impact user satisfaction of and user interest in the experience with health websites?

Methods

Design Overview

This study was a 2×2 (modality interactivity: high vs low; message interactivity: high vs low) online experiment to evaluate effects of modality interactivity and message interactivity on user experience, knowledge, attitudes, and behavioral intentions.

Participants and Sample

Participants were recruited from Amazon Mechanical Turk, an online crowdsourcing marketplace for tasks, which are posted on the platform. The platform can provide a more diverse and valid sample according to previous studies [54,55]. Participants who had a preexisting account on Mechanical Turk took part in the study. Two attention checks were added in the survey. One is the commonly used instructional manipulation checks, which can increase the statistical power and reliability of a data set [56]. In this attention check, participants must not click on anything and directly go to the next page. The other attention check was a statement at the end of a set of questions asking the participants to choose one specific number to make sure they were reading the questions. Data from individuals who did not pass these 2 checks were not included in the data set. We also had manipulation checks for each type of interactivity. For the modality interactivity, participants were asked if they dragged the slider bar to view the change of the pictures. If the respondents answered “no” to the question, they were dropped from the data. For message interactivity, participants were asked if they clicked on the plus icon to view more information about sun protection. The respondents who answered “no” were dropped as well. Additionally, perceived interactivity was used to evaluate the interactivity level in each condition.

Procedure

An online questionnaire (Qualtrics XM) was used to collect data from the respondents. All the independent variables were between-subject factors. Participants gave consent to participate in this study and were randomly assigned to 1 of 4 experimental conditions. Each participant was provided with a link and asked to explore the websites as much as possible. They were instructed to read all the information on the website and click on as many links and buttons as possible. After browsing the website, each participant completed a questionnaire regarding the website, their perceptions about skin cancer, and sun protection and provided demographic information. At the beginning of the survey, a prompt asking them if they had browsed the website. Such reminders have been found to be one way to increase the viewing of content. The incentive for each participant was US $0.50. The study was approved by Washington State University's institutional review board.

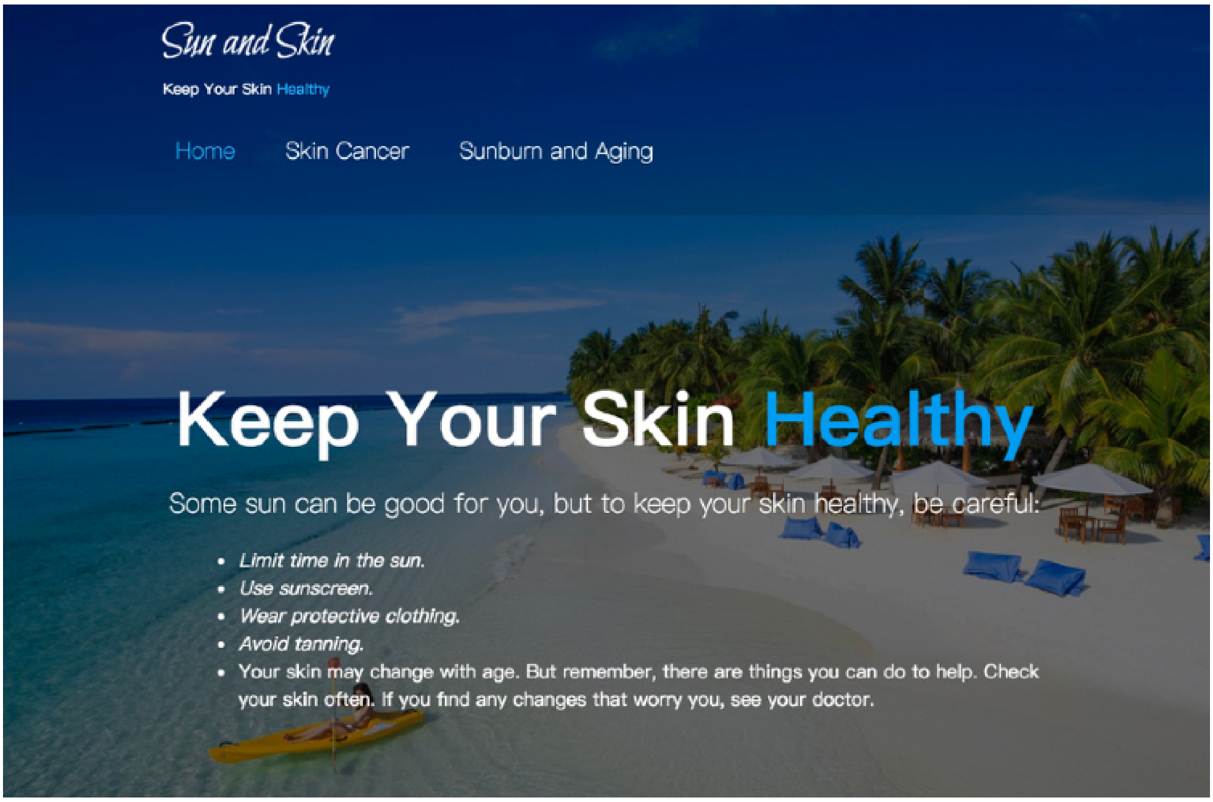
Experimental Treatment Conditions

Four websites were built for this research project. All 4 websites had the same webpage layout and health content. They only differed in terms of interactive features. The websites’ title was “Sun and Skin” with sections on the webpage including one about skin cancer and the other one about sunburn and aging (Figure 1).

Message interactivity was manipulated in the Skin Cancer section of the websites. High message interactivity condition had a clicking function, which the participants needed to click on the bars to get further information (Figure 2). The low message interactivity condition did not have the clicking function, and participants could read the information by scrolling down the page. This is consistent with previous manipulations of message interactivity [24].

Modality interactivity was manipulated in the Sunburn and Aging section of the websites. The high modality interactivity condition had a slider bar feature, which the participants could slide from left to right to view the process of aging (Figure 3). In the low modality interactivity condition, 2 pictures of aging were directly placed on the webpage. This is consistent with previous manipulations of modality interactivity [24].

Figure 1. Screenshot of the website homepage.

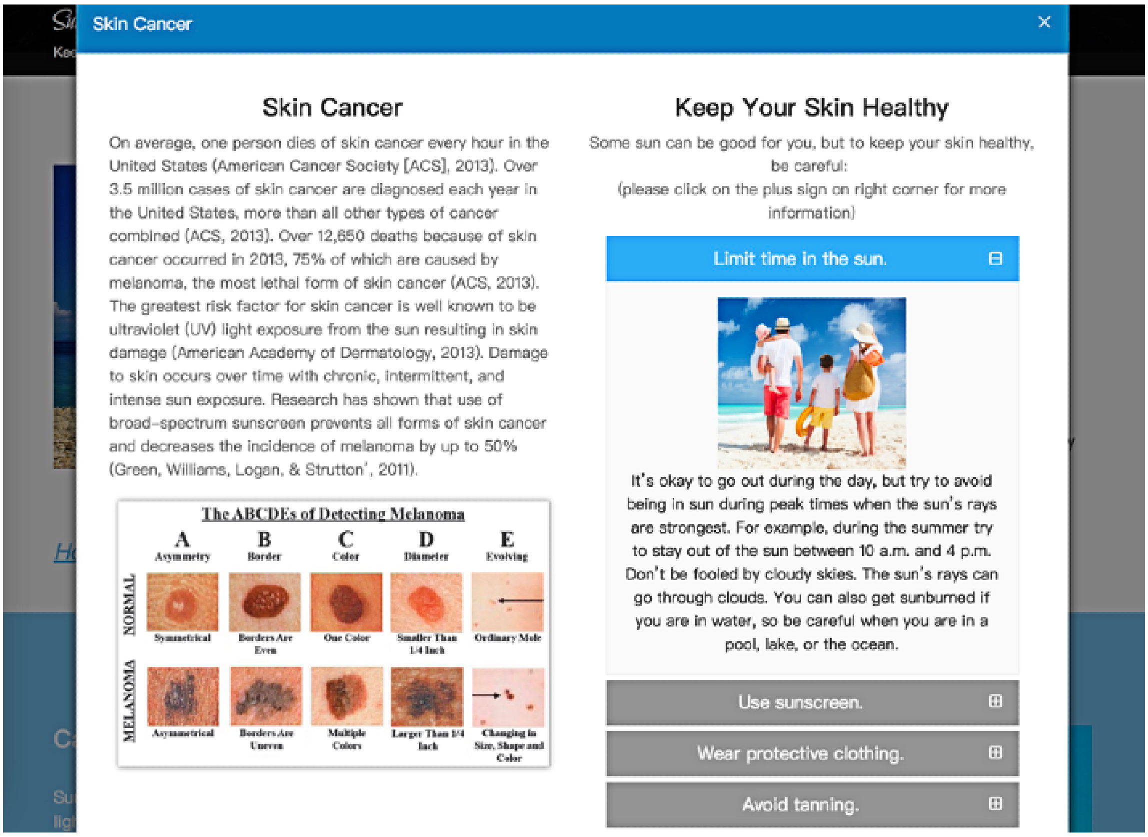


Keep Your Skin Healthy

Some sun can be good for you, but to keep your skin healthy, be careful:

- *Limit time in the sun.*
- *Use sunscreen.*
- *Wear protective clothing.*
- *Avoid tanning.*
- Your skin may change with age. But remember, there are things you can do to help. Check your skin often. If you find any changes that worry you, see your doctor.

Figure 2. Example of the high message interactivity webpage.



Skin Cancer

On average, one person dies of skin cancer every hour in the United States (American Cancer Society [ACS], 2013). Over 3.5 million cases of skin cancer are diagnosed each year in the United States, more than all other types of cancer combined (ACS, 2013). Over 12,650 deaths because of skin cancer occurred in 2013, 75% of which are caused by melanoma, the most lethal form of skin cancer (ACS, 2013). The greatest risk factor for skin cancer is well known to be ultraviolet (UV) light exposure from the sun resulting in skin damage (American Academy of Dermatology, 2013). Damage to skin occurs over time with chronic, intermittent, and intense sun exposure. Research has shown that use of broad-spectrum sunscreen prevents all forms of skin cancer and decreases the incidence of melanoma by up to 50% (Green, Williams, Logan, & Strutton', 2011).

Keep Your Skin Healthy

Some sun can be good for you, but to keep your skin healthy, be careful:
(please click on the plus sign on right corner for more information)

It's okay to go out during the day, but try to avoid being in sun during peak times when the sun's rays are strongest. For example, during the summer try to stay out of the sun between 10 a.m. and 4 p.m. Don't be fooled by cloudy skies. The sun's rays can go through clouds. You can also get sunburned if you are in water, so be careful when you are in a pool, lake, or the ocean.

Figure 3. Example of high modality interactivity webpage.

Photoaging describes the clinical and histologic skin changes induced by sun exposure. Affected skin loses elasticity and appears dry, wrinkled, and patchily pigmented and often has dilated superficial blood vessels and actinic keratoses.

Ultraviolet (UV) A and B components of solar radiation are implicated in photoaging of the skin. Apart from unwanted cosmetic effects, photoaging is a strong risk factor for skin cancer. Ultraviolet radiation damages nucleic acids and proteins in epidermal cells directly and through reactive oxygen species, resulting in impaired collagen and elastin homeostasis, local immune suppression, altered differentiation of keratinocytes, and ultimately tumor development.

A long-term clinical trial that took more than 2 years in Australia proved that regular sunscreen use by young and middle-aged adults younger than 55 years can retard skin aging.

An estimated 90 percent of skin aging is caused by the sun. People who use sunscreen with an SPF of 15 or higher daily show 24 percent less skin aging than those who do not use sunscreen daily.

It's as easy to get sunburned in winter as in summer. Snow reflects the sun's rays. Too much exposure to UVB rays can lead to sunburn. UVA rays can travel more deeply into the skin than UVB rays, but both can affect your skin's health. When UV rays enter skin cells, they upset delicate processes that affect the skin's growth and appearance.

Over time, exposure to these rays can make the skin less elastic. Skin may even become thickened and leathery, wrinkled, or thinned like tissue paper. "The more sun exposure you have, the earlier your skin ages," says Dr. Barnett S. Kramer, a cancer prevention expert at NIH.

Measures

Manipulation Checks

Perceived interactivity was measured by 3 items adapted from Kalyanaraman and Sundar [5], asking participants to indicate how interactive the website was, if the website allowed them to perform a lot of actions, and if the website allowed them to access information in a variety of ways. Additionally, a manipulation check item was asked for each of the interactivity manipulation types to assess exposure to the stimuli. To assess message interactivity, participants were asked "Did you click on the link to view more information?" (yes or no response) and "How interactive did you feel the site was when using the drop-down button/menu?" (scale of 1-7, where 1=not at all and 7=extremely). Modality interactivity was assessed with "Did you drag the slider bar to view more information?" (yes or no response) and "How interactive did you feel the site was while dragging the slider to view the picture?" (scale of 1-7, where 1=not at all and 7=extremely).

Dependent Variables

Behavioral intention to use sunscreen was measured by 3 items (using a 7-point Likert scale) items adapted from Kahlor [57] including "I plan to use sunscreen in the future," "I will try to use sunscreen in the future," and "I intend to use sunscreen in the near future" (Cronbach α=.93).

User satisfaction of the experience was measured 5 items (using a 7-point Likert scale) such as "I am totally satisfied with my interaction with this health website" and "This site was very careful in considering my health information needs" (Cronbach α=.76).

User interest in the experience was measured by 3 items (using a 7- point Likert scale) such as "The interaction with this health website was interesting" and "I am interested in the health information presented on the website" (Cronbach α=.76).

Mediating Variables

Attitudes toward the health websites were measured by statements asking the respondents on a 7-point scale whether they felt that the website they just viewed was exciting or not exciting, high quality or low quality, fun or not fun, cool or not cool, imaginative or not imaginative, and entertaining or not entertaining (Cronbach α=.87).

Knowledge was measured from 2 aspects, the objective memory recall and the perception of knowledge increase. Perceived knowledge was measured by 4 items (using a 7-point Likert-scale ranging from 1=strongly disagree to 7=strongly agree) asking how knowledgeable the respondents felt after exploring the websites [20], such as "I feel very knowledgeable about skin cancer" and "I feel very confident about my ability to tell the disadvantages of sun exposure" (Cronbach α=.89). Objective memory recall was measured by 3 questions asking whether they thought the following statements were correct based on what they had read. The statements were retrieved from the information on the website and included "People don't need to use sunscreen in winter," "Tanning pills can help protect skin when you use tanning beds," and "Sunscreen with an SPF of 10 is enough for people doing outdoors activities." Correct answers were coded as 1 and incorrect answers were coded as 0. The number of correct answers was integrated as the final scores for the objective memory recall.

Statistical Analysis

General linear model analyses were used to test the effects of the 2 independent variables (modality interactivity and message interactivity) on the dependent variables. To test the mediating relationships, PROCESS macro (version 3.3) was conducted using SPSS (version 25.0; IBM Corp). Age, gender, race (White vs non-White), and education were controlled as covariates in all analyses. $P<.05$ was used to determine the statistical significance level. Prior to the analysis of the data, all manipulation check items were assessed. Data from participants who did not pass were excluded from analysis. Independent sample *t* tests (2-tailed) were used to check if participants in high interactivity conditions had higher scores on the second set of manipulation check questions.

Results

Manipulation Checks

Participants in the high modality condition (mean 4.67, SD 1.55) scored higher on perceived interactivity than participants in the low modality condition (mean 4.22, SD 1.50; t_{291}=2.58, P=.02); participants in the high message condition (mean 4.60, SD 1.48) scored higher on perceived interactivity than participants in low message condition (mean 4.30, SD 1.60; t_{291}=1.18, P=.03).

User Statistics

Of 316 participants who responded to the survey, data from 293 participants were included in the analysis (participants who were missing data on variables under study, n=7; participants who did not pass manipulation checks, n=16). Demographic characteristics of the sample are shown in Table 1.

Table 1. Demographics of sample.

Variable	Value
Age (years), mean (SD)	35.97 (12.02)
Gender, n (%)	
Male	133 (44.6)
Female	160 (53.7)
Race/ethnicity, n (%)	
Caucasian/White	217 (74.1)
African American/Black	28 (9.5)
Hispanic or Latino	15 (5.1)
Asian/Pacific Islander	29 (9.9)
Other	4 (1.4)
Family income in last year (US $), n (%)	
≤$20,000	35 (11.9)
$20,001-$50,000	111 (37.9)
$50,001-$70,000	52 (17.7)
$70,001-$100,000	61 (20.8)
$100,001-$150,000	23 (7.8)
≥$150,000	11 (3.8)
Education, n (%)	
High school degree or less	24 (8.2)
Some college	87 (29.7)
College degree	120 (41.0)
Some graduate school	17 (5.8)
Graduate degree	45 (15.4)

Modality Interactivity

The main effect of modality interactivity on attitudes toward health websites was significant ($F_{1,283}$=4.02, P=.045, η^2=.014). Participants in the high modality interactivity condition (mean 4.64, SE 0.12) scored higher on attitudes toward health websites than participants in the low modality condition (mean 4.30, SE 0.12). Similarly, the main effect of modality interactivity on the intention to use sunscreen was also significant ($F_{1,283}$=10.59, P=.001, η^2=.036). Participants who experienced high modality websites (mean 5.70, SE 0.12) scored higher on intention to use sunscreen than participants who explored the low modality websites (mean 5.17, SE 0.12). Thus, hypothesis 1 was supported.

Message Interactivity

The main effect of message interactivity on perceived knowledge was significant ($F_{1,283}$=12.08, P=.001, η^2=.041). Participants in the high message interactivity condition (mean 4.98, SE 0.10) scored higher on perceived knowledge than participants in the low message condition (mean 4.47, SE 0.11). However, the main effect of message interactivity on objective memory recall was not significant. Thus, the first part of hypothesis 2 was partially supported.

The main effect of message interactivity on intention to use sunscreen was significant ($F_{1,283}$=17.02, P<.001, η^2=.057). However, participants in the low message interactivity condition (mean 5.78, SE 0.12) reported higher intention to use sun protection than participants in high message interactivity condition (mean 1.10, SE 0.12). Thus, the second part of hypothesis 2 was supported.

Interaction Effects

No interaction effect was found for attitudes toward health website, perceived knowledge, or intention to use sunscreen. However, the interaction effect of modality interactivity and message interactivity on objective memory recall was significant ($F_{1,283}$=12.90, P<.001, η^2=.043), answering the first research question. Participants in high modality and high message interactivity group (mean 0.61, SE 0.08) scored higher on objective memory recall than participants in high modality and low message interactivity group (mean 0.20, SE 0.08).

Mediation Analysis

According to the results of mediation analyses, there was a significant indirect effect of modality interactivity on intention to use sun protection via its prior effect on attitudes toward health websites ($B_{indirect}$=0.072, SE.044, 95% CI 0.003 to 0.173) (Figure 4). Thus, hypothesis 3 was supported.

In addition, there was a significant indirect effect of message interactivity on intention to use sun protection via its prior effect on perceived knowledge ($B_{indirect}$=0.082, SE 0.046, 95% CI 0.009 to 0.187) (Figure 5). However, objective knowledge gain was not a mediator of message interactivity on intention to use sun protection. Thus, hypothesis 4 was partially supported.

Figure 4. Mediation model of modality interactivity on intention through attitudes, including effects of control variables (age, sex, race, and education, which are not displayed). **P*<.05; ***P*<.01; ****P*<.001.

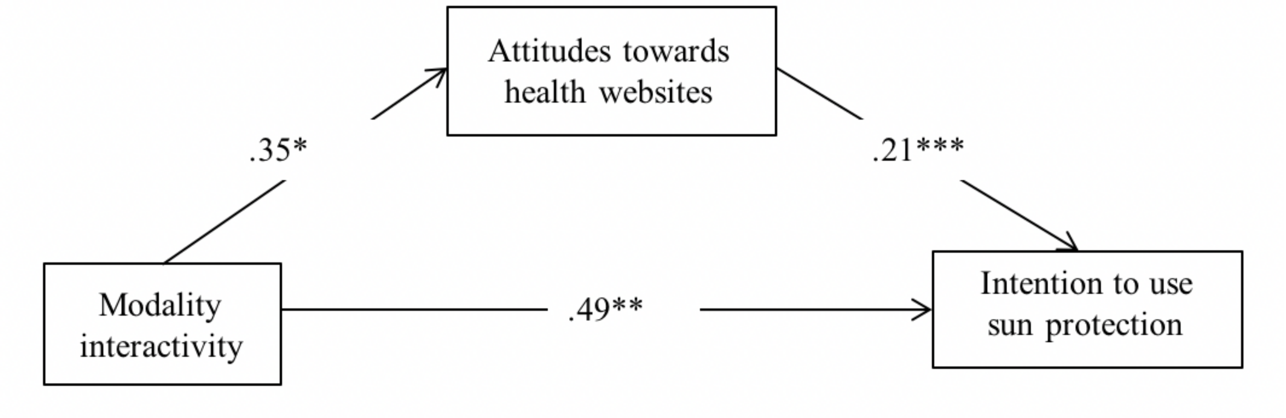
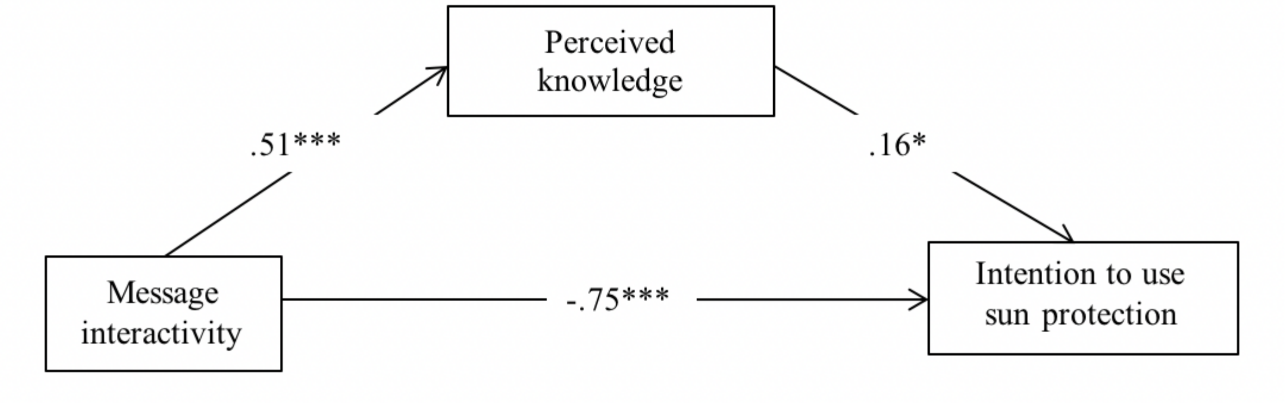


Figure 5. Mediation model of message interactivity on intention through attitudes, including effects of control variables (age, sex, race, and education, which are not displayed). **P*<.05; ***P*<.01; ****P*<.001.



User Satisfaction and User Interest

According to the results of two 2-way analysis of covariance, there was no main effect of modality interactivity or message interactivity on user satisfaction; however, the interaction effect of modality interactivity and message interactivity on user satisfaction was significant ($F_{1,283}$=4.52, *P*=.023, η^2=.01). Participants in the high modality interactivity and high message interactivity condition (mean 4.70, SE 0.14) felt more satisfied compared to participants in the low modality interactivity condition and high message interactivity condition (mean 4.34, SE 0.14) (Figure 6). The main effect of message interactivity on user interest was also significant ($F_{1,283}$=4.08, *P*=.044, η^2=.01). Participants in the low message interactivity condition (mean 5.68, SE 0.16) were more interested in the health website user experience than participants in the high message interactivity condition (mean 5.06, SE 0.16) (Figure 7), answering the second research question.

Figure 6. Modality interactivity and message interactivity effect on user satisfaction.

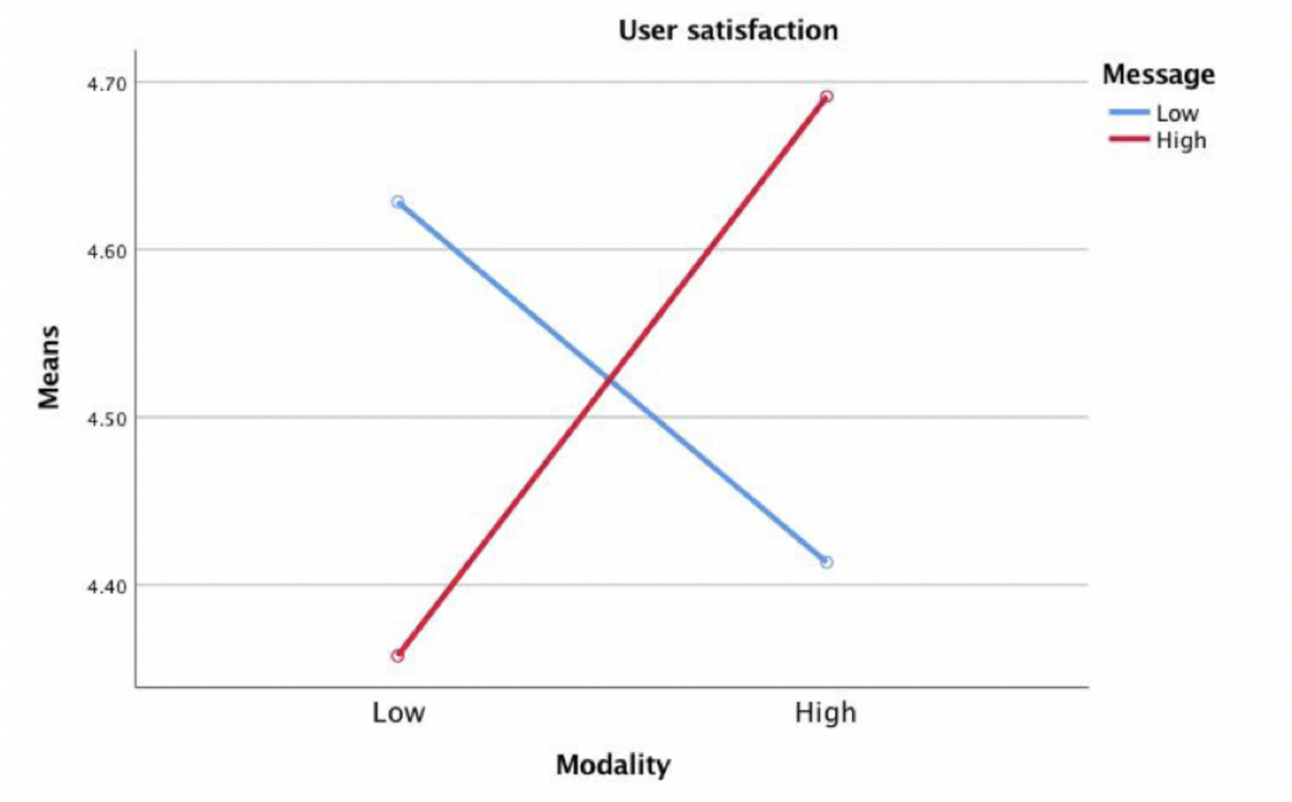
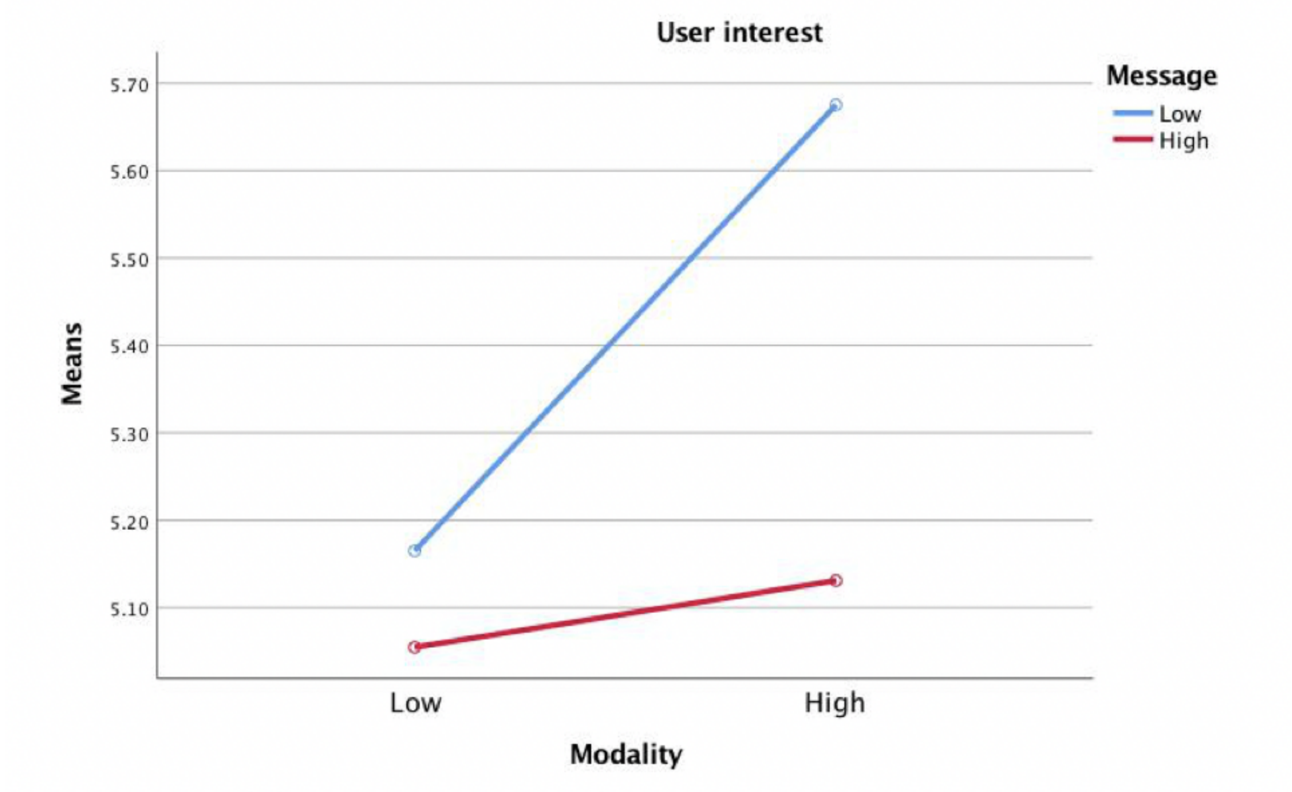


Figure 7. Modality interactivity and message interactivity effect on user interest.



Discussion

General

This study aimed to examine both the main effects and interaction effects of 2 types of interactivity presented on health websites regarding skin cancer. This study tries to illustrate the distinct effect of different types of interactivity in an empirical study and extends the literature by investigating skin cancer–related perceptions, interaction effects of different types of interactivity, and how interactivity impact user experience with health websites. This study has both theoretical and empirical implications.

Behavioral Intentions

Different types of interactivity influenced behavioral intention to use sunscreen differently through different mediators. Message interactivity had a direct negative effect on the intention to use sunscreen. The reason could be that message interactivity takes more cognitive effort or the process of information acceptance was not favorable [6]. Therefore, individuals who only experienced message interactivity using the skin cancer website without additional modality interaction had lower intention to use sunscreen in the future. Modality interactivity had a direct positive effect on intention to use sunscreen. Individuals who experienced greater modality features on the website were more likely to use sunscreen in the future. Therefore, health researchers and health campaign designers can decrease message interactivity and increase modality interactivity on the health website to achieve the goal of promoting behavioral intention to use sunscreen.

Attitudes and Knowledge

Attitudes and perceived knowledge were found to mediate the effects of different interactivity on intention to use sunscreen. Individuals who experienced a high level of modality features on the health websites tended to have more favorable attitudes toward the health websites, which includes their evaluation of website design, interaction with the features, and user experience of the whole website.

One contribution of this study is that both perceived knowledge and actual knowledge gain were tested to see if there was a difference between how these 2 types of knowledge were impacted by interactivity and how they influenced behavioral intention. Message interactivity had a positive effect on perceived knowledge. Participants in the high message interactivity condition felt more knowledgeable than participants in the low message interactivity condition. But there was no main effect of interactivity on actual knowledge gain (memory recall). Although participants perceived that their knowledge about skin cancer increased, their actual knowledge gain was not significantly affected by interactivity.

Skin cancer is the most preventable cancer compared to other forms of cancer and prevention of skin cancer has tremendous potential to save lives [35]. Preventive behaviors such as using sun protection are also relatively easy to perform for the general public. Lustria [21] found that the interactivity level of a website had a significant effect on the comprehension of the skin cancer content and attitudes toward health websites. This research extended the findings of previous research [21,24] testing interactivity on skin cancer–related outcomes and distinguished effects of different types of interactivity. Results of this study provide insights for the design of future skin cancer interventions, especially those using the web-based platforms [21,42,43].

These results also provide a direction for other health topics. Attitudes toward the website influenced behavioral intentions to perform actions related to the health topic, which should draw attention from health practitioners to website design. If the primary goal is to achieve health behavior change, health practitioners should focus on how to increase favorable attitudes toward health websites when designing health websites.

User Experience

Previous digital health interventions have not focused on user experience with the media system or the program; however, the user experience may directly impact users' absorption of the information and future behaviors [58]. Our study found support that interaction effect of modality interactivity and message interactivity had a positive influence on user satisfaction whereas participants experienced high modality interactivity and high message interactivity reported highest satisfaction scores on their interaction with the health websites. Additionally, participants in the low message interactivity condition scored the highest on the interest in their experience with the health websites. Future studies should investigate how user experience impact behavioral outcomes related to health.

Theoretical Contribution and Implications in Health Interventions

In computer-mediated communication or human-computer interaction fields of study, scholars tend to use their own definitions or dimensions of interactivity in their studies [59]. Therefore, there is not a unified definition of interactivity. We used the MAIN model [6] as a theoretical background to conceptualize different types of interactivity and to map the relationships between interactivity, mediators, and behavioral outcomes. Previous studies have examined some interactivity models in the advertising field (eg, a dual-process model of interactivity effects [47]) or focused on building a moderation model of technological attributes (eg, mediated moderation model of interactivity [60]). However, there are not many studies explicitly assessing the influences of different interactivity in eHealth or mobile health (mHealth) interventions [59]. Employing interactivity in health interventions still requires additional empirical studies. Our findings lend support to most of the theoretically constructed hypotheses and demonstrates that different types of interactivity have different influences on health behavioral intention through different mediators. The technological affordances on the media systems need to be carefully defined and applied in eHealth and mHealth interventions.

This study also has empirical implications in designing health interventions or health campaigns. Previous studies have used interactivity as a general concept and have not differentiated among various technological functions, such as hyperlinks and 3D rotation function (eg, hyperlinks [61]; zoom-in and zoom-out, pan and rotate [20]; navigation tools and hyperlinks [21]). This study distinguished message interactivity and modality interactivity and operationalized these 2 types of interactivity with different technological features in an experimental study and explicitly examined the distinct effects of different types of interactivity on a health website, which has filled the previous literature gap in interactivity research and has added methodological implications to empirical eHealth research. Future health interventions using interactive websites or any interactive futures on the interface could apply the findings from this study to achieve desired goals.

Limitations and Future Studies

Like many studies, this study has some limitations. First of all, Mechanical Turk data were used in this study. Mechanical Turk data provide more socioeconomically and demographically diverse samples than those obtained from college and traditional internet sampling [54]. However, as an online experiment, we had less control during the data collection process. The process of how viewers browsed the websites was not clear. Additionally, the lack of control could also influence participants' experience with the website. While participants were directed to use the features of the website, we do not have data related to the length of time spent on each webpage or the number of interactivity features accessed. While only participants who passed the manipulation check remained in the study, there could also be differences based on the level of interaction with which participants engaged with the website, which we were unable to examine in this study. However, the use of an online experiment mimics the real situation of how people may view health websites at home or other places in their daily life instead of a research lab. This may lend additional external validity to the study. Given the rapid development of new technologies on websites, future studies should also employ new interactive features and establish user-centered websites with more professional functions.

Participants who had been diagnosed with skin cancer or had relatives who had skin cancer might be more engaged or be impacted more by skin cancer interventions. Future studies should distinguish the effects of different experiences with skin cancer. Additionally, the outcome variables in this study were attitudes and behavioral intentions instead of actual behaviors. People may overreport their behavioral intentions to perform healthy activities for social desirability. Therefore, future studies should aim at measuring actual behavior change such as sun protection behaviors to investigate the behaviors in real-world situations. A longitudinal study is needed to evaluate how interactive features work in changing people's health behaviors, which might be the ultimate goal of health campaigns and interventions.

Conclusions

This study, which used a 2×2 experimental design to assess 2 different types of interactivity, has contributions related to designing effective health websites for health interventions. Different types of interactivity along with attitudes, knowledge, and behavioral intention were examined to map the mediating relationships between independent and dependent variables. Both modality interactivity and message interactivity had direct positive effects on behavioral intention to use sunscreen. Modality interactivity also had an indirect positive effect on behavioral intention through attitudes toward the website. Message interactivity had an indirect effect on behavioral intention to use sunscreen via perceived knowledge. To design a health intervention or campaign in the digital age, health researchers and practitioners could employ interactive features in their designs accordingly. This study has important insights for health practitioners who have different aims when designing health websites for eHealth and mHealth interventions.

Conflicts of Interest

None declared.

References

1. Liu Y, Shrum LJ. What is interactivity and is it always such a good thing? implications of definition, person, and situation for the influence of interactivity on advertising effectiveness. Journal of Advertising 2002 Dec;31(4):53-64. [doi: 10.1080/00913367.2002.10673685]
2. Rice R. The New Media: Communication, Research, and Technology. Beverly Hills, CA: SAGE Publications, Inc; 1984.
3. Cassell MC, Jackson C, Cheuvront B. Health communication on the internet: an effective channel for health behavior change? J Health Commun 1998 Jan;3(1):71-79. [doi: 10.1080/108107398127517] [Medline: 10947375]
4. Kim H, Stout PA. The effects of interactivity on information processing and attitude change: implications for mental health stigma. Health Commun 2010 Mar 26;25(2):142-154. [doi: 10.1080/10410230903544936] [Medline: 20390680]
5. Kalyanaraman S, Sundar S. The psychological appeal of personalized content in web portals: does customization affect attitudes and behavior? Journal of Communication 2006;56(1):110-132. [doi: 10.1111/j.1460-2466.2006.00006.x]
6. Sundar S. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Miriam J. M, Andrew J. F, editors. Digital Media, Youth, and Credibility. Cambridge, MA: The MIT Press; 2008:73-100.
7. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
8. Harvey-Berino J, Pintauro S, Buzzell P, DiGiulio M, Casey Gold B, Moldovan C, et al. Does using the internet facilitate the maintenance of weight loss? Int J Obes Relat Metab Disord 2002 Sep 21;26(9):1254-1260. [doi: 10.1038/sj.ijo.0802051] [Medline: 12187404]

9. Homer C, Susskind O, Alpert HR, Owusu MS, Schneider L, Rappaport LA, et al. An evaluation of an innovative multimedia educational software program for asthma management: report of a randomized, controlled trial. Pediatrics 2000 Jul;106(1 Pt 2):210-215. [Medline: 10888694]
10. Levine D. Using technology, new media, and mobile for sexual and reproductive health. Sex Res Soc Policy 2011 Feb 25;8(1):18-26. [doi: 10.1007/s13178-011-0040-7]
11. Marshall AL, Leslie ER, Bauman AE, Marcus BH, Owen N. Print versus website physical activity programs. Am J Prev Med 2003 Aug;25(2):88-94. [doi: 10.1016/s0749-3797(03)00111-9]
12. Oenema A, Brug J, Lechner L. Web-based tailored nutrition education: results of a randomized controlled trial. Health Educ Res 2001 Dec;16(6):647-660. [doi: 10.1093/her/16.6.647] [Medline: 11780705]
13. Ritterband LM, Cox DJ, Walker LS, Kovatchev B, McKnight L, Patel K, et al. An internet intervention as adjunctive therapy for pediatric encopresis. J Consult Clin Psychol 2003;71(5):910-917. [doi: 10.1037/0022-006x.71.5.910]
14. Health online 2013. Pew Research Center. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2017-06-06]
15. Baker DB. Patient-centered healthcare. Disease Management and Health Outcomes 2001;9(8):411-420. [doi: 10.2165/00115677-200109080-00001]
16. Eng TR. eHealth research and evaluation: challenges and opportunities. J Health Commun 2002 Jul;7(4):267-272. [doi: 10.1080/10810730290001747] [Medline: 12356287]
17. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
18. Rafaeli S. From new media to communication. In: Sage Annual Review of Communication Research: Advancing Communication Science. Beverly Hills, CA: SAGE Publications, Inc; 1988:110-134.
19. Atkinson NL, Gold RS. The promise and challenge of eHealth interventions. Am J Health Behav 2002 Nov 01;26(6):494-503. [doi: 10.5993/ajhb.26.6.10] [Medline: 12437024]
20. Lu Y, Kim Y, Dou X, Kumar S. Promote physical activity among college students: using media richness and interactivity in web design. Comput Human Behav 2014 Dec;41:40-50. [doi: 10.1016/j.chb.2014.08.012]
21. Lustria MLA. Can interactivity make a difference? effects of interactivity on the comprehension of and attitudes toward online health content. J Am Soc Inf Sci 2007 Apr;58(6):766-776. [doi: 10.1002/asi.20557]
22. Jaffe JM. Media interactivity and self-efficacy: an examination of hypermedia first aid instruction. J Health Commun 1997 Nov;2(4):235-252. [doi: 10.1080/108107397127572]
23. Neuhauser L, Kreps GL. Rethinking communication in the e-health era. J Health Psychol 2003 Jan;8(1):7-23. [doi: 10.1177/1359105303008001426] [Medline: 22113897]
24. Oh J, Sundar SS. How does interactivity persuade? an experimental test of interactivity on cognitive absorption, elaboration, and attitude. J Commun 2015 Feb 27;65(2):213-236. [doi: 10.1111/jcom.12147]
25. Oh J, Sundar S. User engagement with interactive media: a communication perspective. In: In Why Engagement Matters. Basel, Switzerland: Springer International Publishing; 2016:177-198.
26. Willoughby JF, L'Engle KL. Influence of perceived interactivity of a sexual health text message service on young people's attitudes, satisfaction and repeat use. Health Educ Res 2015 Dec 20;30(6):996-1003. [doi: 10.1093/her/cyv056] [Medline: 26590247]
27. Camerini L, Schulz PJ. Effects of functional interactivity on patients' knowledge, empowerment, and health outcomes: an experimental model-driven evaluation of a web-based intervention. J Med Internet Res 2012 Jul 18;14(4):e105 [FREE Full text] [doi: 10.2196/jmir.1953] [Medline: 22810046]
28. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
29. Bucy EP. Interactivity in society: locating an elusive concept. The Information Society 2004 Nov;20(5):373-383. [doi: 10.1080/01972240490508063]
30. Sundar S, Marathe S. Personalization versus customization: the importance of agency, privacy, and power usage. Human Communication Research 2010 Jul 01;36(3):298-322. [doi: 10.1111/j.1468-2958.2010.01377.x]
31. Guy GP, Machlin SR, Ekwueme DU, Yabroff KR. Prevalence and costs of skin cancer treatment in the U.S., 2002-2006 and 2007-2011. Am J Prev Med 2015 Feb;48(2):183-187 [FREE Full text] [doi: 10.1016/j.amepre.2014.08.036] [Medline: 25442229]
32. Stern RS. Prevalence of a history of skin cancer in 2007: results of an incidence-based model. Arch Dermatol 2010 Mar 01;146(3):279-282. [doi: 10.1001/archdermatol.2010.4] [Medline: 20231498]
33. Armstrong AW, Watson AJ, Makredes M, Frangos JE, Kimball AB, Kvedar JC. Text-message reminders to improve sunscreen use: a randomized, controlled trial using electronic monitoring. Arch Dermatol 2009 Nov 01;145(11):1230-1236. [doi: 10.1001/archdermatol.2009.269] [Medline: 19917951]
34. Siegel R, Miller K, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66(1):7-30 [FREE Full text] [doi: 10.3322/caac.21332] [Medline: 26742998]

35. Cokkinides V, Albano J, Samuels A, Ward M, Thum J. Cancer Facts and Figures, 2005. Atlanta, Georgia: American Cancer Society, Inc; 2005.
36. Hughes BR, Altman DG, Newton JA. Melanoma and skin cancer: evaluation of a health education programme for secondary schools. Br J Dermatol 1993 Apr;128(4):412-417. [doi: 10.1111/j.1365-2133.1993.tb00201.x] [Medline: 8494755]
37. Kamin C, O'Neill P, Ahearn M. Developing and evaluating a cancer prevention teaching module for secondary education: Project SAFETY (Sun Awareness for Educating Today's Youth). J Cancer Educ 1993;8(4):313-318. [doi: 10.1080/08858199309528248] [Medline: 8186082]
38. Mermelstein RJ, Riesenberg LA. Changing knowledge and attitudes about skin cancer risk factors in adolescents. Health Psychology 1992;11(6):371-376. [doi: 10.1037/0278-6133.11.6.371]
39. Hingle MD, Snyder AL, McKenzie NE, Thomson CA, Logan RA, Ellison EA, et al. Effects of a short messaging service-based skin cancer prevention campaign in adolescents. Am J Prev Med 2014 Nov;47(5):617-623 [FREE Full text] [doi: 10.1016/j.amepre.2014.06.014] [Medline: 25053602]
40. Janda M, Youl P, Marshall A, Soyer H, Baade P. The HealthyTexts study: a randomized controlled trial to improve skin cancer prevention behaviors among young people. Contemp Clin Trials 2013 May;35(1):159-167. [doi: 10.1016/j.cct.2013.03.009] [Medline: 23557730]
41. Youl PH, Soyer HP, Baade PD, Marshall AL, Finch L, Janda M. Can skin cancer prevention and early detection be improved via mobile phone text messaging? a randomised, attention control trial. Prev Med 2015 Feb;71:50-56. [doi: 10.1016/j.ypmed.2014.12.009] [Medline: 25524612]
42. Bernhardt J. Tailoring messages and design in a Web-based skin cancer prevention intervention. The International Electronic Journal of Health Education; 2001;4:290-297.
43. Heckman C, Darlow S, Munshi T, Caruso C, Ritterband L, Raivitch S, et al. Development of an internet intervention to address behaviors associated with skin cancer risk among young adults. Internet Interv 2015 Sep;2(3):340-350 [FREE Full text] [doi: 10.1016/j.invent.2015.04.003] [Medline: 26640776]
44. Sundar S, Xu Q, Bellur S. Designing interactivity in media interfaces: a communications perspective. 2010 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 10; Atlanta, Georgia p. 2247-2256. [doi: 10.1145/1753326.1753666]
45. Sundar SS. Multimedia effects on processing and perception of online news: a study of picture, audio, and video downloads. Journalism & Mass Communication Quarterly 2016 Jun 25;77(3):480-499. [doi: 10.1177/107769900007700302]
46. Reeves B, Nass C. Perceptual user interfaces: perceptual bandwidth. Commun. ACM 2000 Mar;43(3):65-70. [doi: 10.1145/330534.330542]
47. Liu Y, Shrum LJ. A dual-process model of interactivity effects. Journal of Advertising 2009 Jul;38(2):53-68. [doi: 10.2753/joa0091-3367380204]
48. Sundar SS, Kalyanaraman S, Brown J. Explicating web site interactivity. Communication Research 2016 Jun 30;30(1):30-59. [doi: 10.1177/0093650202239025]
49. Niu Z, Jeong D, Coups E, Stapleton J. An experimental investigation of human presence and mobile technologies on college students? sun protection intentions: between-subjects study. JMIR Mhealth Uhealth 2019 Aug 26;7(8):e13720 [FREE Full text] [doi: 10.2196/13720] [Medline: 31452523]
50. Oviatt S, MacEachern M, Levow G. Predicting hyperarticulate speech during human-computer error resolution. Speech Communication 1998 May;24(2):87-110. [doi: 10.1016/s0167-6393(98)00005-3]
51. Sundar S, Xu Q, Bellur S, Oh J, Jia H. Beyond pointing and clicking: how do newer interaction modalities affect user engagement? InCHI'11 Extended Abstracts on Human Factors in Computing Systems 2011 May 07:1477-1482. [doi: 10.1145/1979742.1979794]
52. Schlosser A. Experiencing products in the virtual world: the role of goal and imagery in influencing attitudes versus purchase intentions. J Consum Res 2003 Sep;30(2):184-198. [doi: 10.1086/376807]
53. Hurling R, Fairley BW, Dias MB. Internet-based exercise intervention systems: are more interactive designs better? Psychology & Health 2006 Dec;21(6):757-772. [doi: 10.1080/14768320600603257]
54. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
55. Casler K, Bickel L, Hackett E. Separate but equal? a comparison of participants and data gathered via Amazon’s MTurk, social media, and face-to-face behavioral testing. Computers in Human Behavior 2013 Nov;29(6):2156-2160. [doi: 10.1016/j.chb.2013.05.009]
56. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: detecting satisficing to increase statistical power. Journal of Experimental Social Psychology 2009 Jul;45(4):867-872. [doi: 10.1016/j.jesp.2009.03.009]
57. Kahlor L. PRISM: a planned risk information seeking model. Health Commun 2010 Jun 28;25(4):345-356. [doi: 10.1080/10410231003775172] [Medline: 20512716]
58. Ye C, Seo D, Desouza KC, Sangareddy SP, Jha S. Influences of IT substitutes and user experience on post-adoption user switching: An empirical investigation. J. Am. Soc. Inf. Sci 2008 Nov;59(13):2115-2132. [doi: 10.1002/asi.20921]
59. Willoughby JF, Niu Z. A review of the influence of interactivity on health-related outcomes and recommendations for future study. iproc 2016 Dec 30;2(1):e41 [FREE Full text] [doi: 10.2196/iproc.6093]

60. Bucy EP, Tao C. The mediated moderation model of interactivity. Media Psychology 2007 May 15;9(3):647-672. [doi: 10.1080/15213260701283269]
61. Chung H, Zhao X. Effects of perceived interactivity on web site preference and memory: role of personal motivation. Journal of Computer-Mediated Communication 2004;10(1):JCMC1017. [doi: 10.1111/j.1083-6101.2004.tb00232.x]

Abbreviations

MAIN: Modality, Agency, Interactivity, and Navigability
mHealth: mobile health

Edited by G Eysenbach; submitted 20.04.20; peer-reviewed by J Nwaogu, C Ang, T Steeb; comments to author 15.06.20; revised version received 14.08.20; accepted 26.10.20; published 13.01.21.

Please cite as:
Niu Z, Willoughby JF, Coups EJ, Stapleton JL
Effects of Website Interactivity on Skin Cancer–Related Intentions and User Experience: Factorial Randomized Experiment
J Med Internet Res 2021;23(1):e18299
URL: http://www.jmir.org/2021/1/e18299/
doi:10.2196/18299
PMID:33439131

Original Paper

Effects of Web-Based Social Connectedness on Older Adults' Depressive Symptoms: A Two-Wave Cross-Lagged Panel Study

Juwon Hwang[1], BA, MA; Catalina L Toma[2], BA, MSc, PhD; Junhan Chen[3], BA, MA; Dhavan V Shah[1], BA, MA, PhD; David Gustafson[4], BSc, MSc, PhD; Marie-Louise Mares[2], BA, MA, PhD

[1]School of Journalism and Mass Communication, University of Wisconsin, Madison, WI, United States
[2]Department of Communication Arts, University of Wisconsin-Madison, Madison, WI, United States
[3]Department of Communication, University of Maryland, College Park, MD, United States
[4]Center for Health Enhancement Systems Studies, University of Wisconsin, Madison, WI, United States

Corresponding Author:
Juwon Hwang, BA, MA
School of Journalism and Mass Communication
University of Wisconsin
5007 Vilas Hall
821 University Ave.
Madison, WI, 53706
United States
Phone: 1 6083324571
Email: jhwang79@wisc.edu

Abstract

Background: Depressive symptoms are the most prevalent mental health concern among older adults (possibly heightened during the COVID-19 pandemic), which raises questions about how such symptoms can be lowered in this population. Existing research shows that *offline* social connectedness is a protective factor against depression in older adults; however, it is unknown whether *web-based* social connectedness can have similar effects.

Objective: This study investigates whether social connectedness on a support website protects older adults against depressive symptoms over the course of a year, above and beyond the protective effect of offline social connectedness. The secondary aim is to determine whether older adults with increased depressive symptoms are more likely to engage in social connectedness on this website. Thus, we examine depressive symptoms as both an outcome and predictor of web-based social connectedness to fully understand the chain of causality among these variables. Finally, we compare web-based social connectedness with offline social connectedness in their ability to lower depressive symptoms among older adults.

Methods: A total of 197 adults aged 65 years or older were given access to a social support website, where they were able to communicate with each other via a discussion forum for a year. Participants' social connectedness on the web-based platform, conceptualized as message production and consumption, was measured using behavioral log data as the number of messages participants wrote and read, respectively, during the first 6 months (t_1) and the following 6 months (t_2) of the study. Participants self-reported their offline social connectedness as the number of people in their support networks, and they reported their depressive symptoms using the Patient Health Questionnaire-8 both at baseline (t_1) and at 12-month follow-up (t_2). To ascertain the flow of causality between these variables, we employed a cross-lagged panel design, in which all variables were measured at t_1 and t_2.

Results: After controlling for the effect of offline support networks at t_1, web-based message consumption at t_1 decreased older adults' depressive symptoms at t_2 (β=−.11; *P*=.02), but web-based message production at t_1 did not impact t_2 depressive symptoms (β=.12; *P*=.34). Web-based message consumption had a larger effect (β=−.11; *P*=.02) than offline support networks (β=−.08; *P*=.03) in reducing older adults' depressive symptoms over time. Higher baseline depressive symptoms did not predict increased web-based message consumption (β=.12; *P*=.36) or production (β=.02; *P*=.43) over time.

Conclusions: The more messages older adults read on the web-based forum for the first 6 months of the study, the less depressed they felt at the 1-year follow-up, above and beyond the availability of offline support networks at baseline. This pinpoints the substantial potential of web-based communication to combat depressive symptoms in this vulnerable population.

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-015-0713-2

(J Med Internet Res 2021;23(1):e21275) doi:10.2196/21275

KEYWORDS

depressive symptoms; older adults; web-based intervention; online social support; patient health questionnaire; longitudinal survey; mobile phone

Introduction

Background

The number of people aged 65 years and older is expected to rise dramatically over the next few decades, accounting for 16.7% of the global population by 2050 [1] and 23.5% of the US population by 2060 [2]. The most prevalent mental health problem among older adults is depression [3], encompassing both major depression and significant depressive symptoms that are below the severity threshold of major depression [4]. Although events such as COVID-19 stay-at-home directives can play a significant role, the multiple changes that occur in late adulthood are largely responsible for depressive symptomology in older adults [5]. For instance, older adults are likely to experience poor health and physical constraints [6] and transition in roles [7], and some older adults experience financial difficulties [6] and social isolation due to retirement and death of friends and family [7]. Both major depression and depressive symptoms cause significant disruption in the daily lives of older adults. Depression is strongly associated with more chronic diseases [4,8], high functional disability and mortality [9], and low quality of life [10]. The prevalence of depressive symptoms among older adults and their harmful consequences underscore the need to understand factors that may guard against the development or continuation of depressive symptoms in this age group.

One such factor is social connectedness [11,12]. Research shows that higher-quality social ties and meaningful social interactions significantly reduce depression in older adults [13-16]. However, the existing literature focuses exclusively on *offline* social connectedness. With internet adoption skyrocketing among this demographic [17], an intriguing possibility is that *web-based* social connectedness has similar protective effects against older adults' depressiveness. In particular, a robust literature shows that online support groups provide options for social connectedness that substantially enhance users' psychological well-being (for a review, see the study by Wright [18]), although research on older adults' participation in these groups is limited and has not yet tackled the issue of depressive symptoms. Therefore, we ask: Does social connectedness via online support groups protect older adults against depressive symptoms above and beyond the well-documented protection provided by social connectedness?

A long-standing difficulty in researching the relationship between social connectedness and depressive symptoms is elucidating the chain of causality. Does social connectedness affect depressive symptoms, or does depressive symptoms affect social connectedness, or both? Indeed, in the realm of *offline* social connectedness, studies show that social connectedness correlates with reduction in depressive symptoms [13,14,16,19], indicating that depressive symptoms can be conceptualized as an outcome of social connectedness. However, other studies show that depressive symptoms correlate with individuals' tendency to withdraw from social situations [20-22], indicating that depressive symptoms can also be conceptualized as an antecedent of social connectedness. Few studies take an integrative approach that considers both causal paths [23,24]. Here, we collect longitudinal data with key variables measures at two time points, which enables us to simultaneously test both causal pathways. Therefore, we are able to make causal inferences about whether web-based social connectedness improves older adults' depressive symptoms over time, and whether older adults are likely to reap the benefits of web-based social connectedness by gravitating toward web-based opportunities for social connection when they experience depressive symptoms.

We begin with a brief review of the benefits of *offline* social connectedness for mitigating depressive symptoms in older adults. Then, we discuss how *web-based* social connectedness may benefit older adults above and beyond the well-established benefits of offline connectedness.

Offline Social Connectedness and Depressive Symptoms Among Older Adults

Considerable research has shown that offline social connectedness is associated with lower depressive symptoms in the older population [13-16]. Importantly, these benefits do not stem from the sheer number of social connections: individuals with large social networks are not immune to loneliness and depression [25,26]. Instead, it is only close, supportive relationships that protect against depressive symptoms [11-13], because they provide individuals with a sense of being cared for, loved, and valued [13,20,27]. Much of this research has relied on cross-sectional designs, which limit causal inference. Nonetheless, a couple of longitudinal studies provide indication of causality. The loss of a spouse [13] and limited in-person contact with family and friends [14] were found to increase depressive symptoms in older adults over time, suggesting that the presence of supportive companions acts as a buffer against depression in this population. In light of this evidence, we expect that the availability of offline support networks will lead to reductions in older adults' depressive symptoms over time.

The Added Benefit of Social Connectedness on Support Websites

Does social connectedness on support websites contribute to older adults' well-being above and beyond offline connectedness? Support websites have become a popular and effective tool for linking individuals with others who share similar life difficulties, yet whom they may never have encountered in everyday life [28]. These websites are rooted in the same principles as those of traditional face-to-face support groups, namely that people's mental and physical health can be bolstered when they perceive their situation as universal rather than unique, when they share information that can help others, and when they receive supportive messages from peers [29-34]. An additional reason for support websites' effectiveness is that they are anonymous and asynchronous, which enables participants to be more forthcoming about difficult, embarrassing, or stigmatizing topics [18,32].

Although the benefits of participation in online social support groups are well documented for those who have health issues such as cancer [30-32], diabetes [35], and HIV or AIDS [36], no research to our knowledge has investigated older adults' social connectedness in these groups in relation to their depressive symptoms. A small body of literature shows that retirement, family, and health ranked among the most popular topics discussed by older adults in web-based communities [37], but that older adults also expressed empathy for other members [38] and developed closely connected subnetworks based on emotional communication [39] in these interaction spaces. This suggests that support websites are conducive to the formation of social bonds between older adults and hold potential for reducing depressive symptoms in this population.

Connecting with peers on support websites can take 2 forms: (1) posting messages and responses to other users' messages (ie, *messageproduction*) and (2) reading other users' postings, thus keeping abreast of others' experiences and insights (ie, *message consumption*). To date, most studies have investigated these 2 forms of social connection separately, finding them both beneficial for participants' well-being [30-32]. Writing messages was shown to alleviate depression, loneliness, pain, and stress among younger individuals [18,40-42]. These benefits accrued because writing in a supportive setting helped participants to reframe their problems in a positive manner and adopt positive coping strategies [43]. Consuming messages by others who were experiencing similar struggles helped participants reduce worry and distress [43,44] because it made them feel less isolated in their struggles [44] and exposed them to different perspectives on a given problem [30]. It bears noting that, contrary to these findings, lurking on social network sites (ie, monitoring others' behaviors without directly communicating with them) has been associated with reductions in well-being [45]. This is likely the case because users of social network sites tend to post glamorized presentations of their lives, thus eliciting envy from viewers [46]. In contrast, users of online support groups tend to show their vulnerabilities and provide encouragement, useful information, and a sense of community to others [47]. Indeed, message consumption on online support groups appears to be helpful even in the absence of message production: participants who read many messages but seldom wrote their own (ie, lurkers [44,48]) experienced increased belonging to the group [48] and reduced isolation [44]. Building on this work, we hypothesize that older adults will also benefit from both web-based message production and consumption within an online support group. These 2 forms of engagement should facilitate the formation of supportive bonds, which in turn should reduce depressive symptoms in this population:

Hypothesis 1. Message production and consumption on a support website will lead to reductions in older adults' depressive symptoms a year later, above and beyond reductions generated by the availability of offline support networks.

The Effects of Web-Based Connectedness Versus Offline Connectedness on Depressive Symptoms

We expect that both web-based and offline social connectedness will serve as protective factors against depressive symptoms in older adults. This raises the question of the relative contribution of these 2 types of connectedness to the amelioration of depressive symptoms.

Research shows that feelings of social connectedness derived from web-based interaction are distinct from those derived from face-to-face interaction, in part because some users connect with different networks online than they do offline. Thus, it is possible to feel connected to web-based networks while disconnected from face-to-face networks and vice-versa [11,49]. In our case, it is unlikely that web-based and offline networks overlap: web-based connections developed on social support websites are likely to be new additions to older adults' offline support networks, rather than duplicates. Nonetheless, both types of connectedness have been shown to be effective in boosting well-being [12,49-51]. As no research has compared the effects of web-based and offline connectedness on alleviating depressive symptoms in general or older adults' depressive symptoms in particular, we pose the following research question:

Research question 1. Is web-based social connectedness or offline social connectedness more potent in reducing depressive symptoms among older adults after 1 year?

Web-Based Connectedness, Depressive Symptoms, and Causal Inference

Finally, let us now consider the issue of chain of causality. As previously hypothesized, we expect psychological benefits to accrue to older adults who connect socially via online support groups. In addition, it is possible that older adults' depressive symptoms affect the extent to which they engage in web-based social connection to begin with. In offline settings, individuals with depression have been shown to withdraw from social situations because of fatigue, lethargy, or diminished feelings of self-worth [20-22]. Depressive symptoms are also predictive of reduced offline connections because depressed individuals' negative self-statements, complaints, and social inadequacy often alienate others [20-22].

To compensate for a lack of offline connectedness, individuals with depressive symptoms might turn to web-based communication, which is substantially less effortful (ie, it can be done without leaving the house) and easier to control because of web-based affordances such as editability and unlimited time to compose messages [52]. Indeed, the social compensation hypothesis [53] predicts that web-based communication is preferred by individuals who experience psychosocial problems that hinder their ability to engage in meaningful face-to-face interactions (eg, anxiety, loneliness, and depression), precisely because it is easy and convenient, and provides an enhanced sense of control over message production and consumption [53,54]. In support of this contention, research shows that participants experiencing more depression tended to increase their engagement with online support groups by posting messages [44] and seeking help from those who experienced similar health issues [29]. Therefore, we predict that older adults with more depressive symptoms are more likely to be engaged in reading and writing messages within an online social support group:

Hypothesis 2. Older adults experiencing more depressive symptoms will engage in greater message production and consumption within an online social support group over time.

Methods

Study Context

The data comes from a clinical trial of an online social support group intervention for older adults named Elder Tree [55]. The intervention consisted of providing participants access to a support website. This website was designed to enable older adults to provide and receive peer support through a web-based discussion forum, with the ultimate goal of helping them maintain their independence and health. The discussion forum enabled older adults to communicate with one another by starting and responding to threads or simply reading each other's messages. Sample messages are provided in Multimedia Appendix 1.

Topics in the forum were not limited to any particular health concerns; rather, they ranged broadly from health to politics or religion. The forum was primarily text-based, although it also supported the uploading of photographs. It was asynchronous, meaning that participants could take all the time they wanted to compose and read messages. Posts could be pseudo-anonymous, meaning that participants were able to express themselves anonymously if they wished, but were also free to disclose their real names and personal photographs. Trained facilitators monitored the conversations to ensure that discussions were supportive and did not contain unchallenged inaccurate or harmful information; however, they did not play an active role in guiding the topics of conversation. Communication took place in a public section, where participants could post messages viewable to all members, as well as in a private section, where participants could post messages only viewable to specific individuals.

Participants and Procedure

Participants were adults aged 65 years and above who had experienced one or more of the following in the last 12 months: a fall, feeling sad or depressed, home-health services, a stay in a skilled nursing facility, an emergency room visit, or admission to the hospital. As our focus was on older adults aging in place, participants were excluded if they were homeless; lived in a hospice center, assisted living facility, or a nursing home; or needed help getting in or out of a bed or a chair. Participants were recruited through senior centers, churches, other community groups, and Aging and Disability Resource Centers. Research coordinators assessed older adults who volunteered for our study; mailed eligible participants the baseline survey; and made home visits to explain the study and obtain consent, collect the baseline survey responses, and explain to participants the expectations of the study based on the group to which they were randomized. Other team members visited participants to give them computers, if needed, and train them on using the website. The training included tutorials on how to use the computer and the internet, as required, but focused on the use of the various features of the support website. Participants were recruited from November 2013 to May 2015, with the intervention ending in November 2016.

Our analyses were conducted on the sample of 197 participants (145/197, 73.6% women; age range: 65-100 years, mean 76.26, SD 7.38) who were given access to the support website. After 12 months, the sample was reduced to 159 due to attrition (80.7% retention rate). It must be noted that the data come from a clinical trial of a support website, where participants were randomly assigned to either access this website or not [55]. In this analysis, only participants who had access to the website were included.

Study Design

This study was set up as a 2-wave cross-lagged panel design, with all variables of interest (ie, depressive symptoms, offline support network size, web-based message consumption, and web-based message production) measured at both time points. The cross-lagged paths allow us to assess causal links between variables over time. As this design allows controlling for variables measured at a previous time point, it rules out the alternative explanations that may occur in cross-sectional studies such as reverse causality and the influence of an auto-regressor [56]. Cross-lagged panel designs are considered the optimal way to understand causality among naturally occurring variables in field settings where experimental procedures are not feasible [57]. Finally, as each person is his or her own control and results can be interpreted as pertaining to the relationship between within-person changes, this design does not require the inclusion of other covariates [58] (see Multimedia Appendix 2 [59,60] for more information on cross-lagged panel designs).

Measures

This study examined longitudinal data with 2 major components: (1) self-reported surveys that were administered at baseline and at the 12-month follow-up and (2) behavioral log data of website use during the study period.

Survey Data

Depressive symptoms were assessed with 8 items of the Patient Health Questionnaire (PHQ) [61] on a 4-point scale (0=not at all, 3=nearly every day) at both time points. Participants rated the degree to which they experienced little interest or pleasure in doing things; feeling down, depressed, or hopeless; and trouble falling or staying asleep, or sleeping too much, for example. The ratings for each item were summed to produce a total score between 0 and 24 points (t1: α=.87; mean 4.42, SD 4.65; t2: α=.87; mean 3.91, SD 4.29). Higher scores indicate higher depressive symptoms. On the basis of the depressive symptom score at baseline, participants were classified into the following depressiveness groups: none to minimal (ie, 0-4 score; 125/197, 63.5% of the participants), mild (ie, 5-9 score; 47/197, 23.8% of the participants), moderate (ie, 10-14 score; 15/197, 7.6% of the participants), moderately severe (ie, 15-19 score; 7/197, 3.6% of the participants), and severe (ie, 20-24 score; 3/197, 1.5% of the participants). Dropout analyses indicated no significant differences in baseline depressive symptoms between the participants in the final sample and those who dropped out between t1 and t2.

The offline support network was assessed with 2 items at both time points. Participants indicated the number of people *you can count on to listen to you when you need to talk and count on you to listen to them when they need to talk*. These items were averaged to produce an overall score (t1: α=.89, mean 4.95, SD 4.80; t2: α=.70, mean 6.21, SD 8.32). Examination of kurtosis values indicated the presence of an outlier case, yielding the ability to skew overall results [62]. The outlier value was winsorized by transforming it with the largest value in observations, excluding the outlier [63,64]. Although outliers are often removed from data, winsorization is recommended for analyses with a small sample size [63].

Behavioral Log Data

Participants' use of the website was captured automatically at an individual keystroke level, as participants used the system. The browser produced a log file with each participant's username, date, and URL of every web page requested from the web server. This enabled us to track the number of messages each participant wrote and/or read in the discussion forum. Recall that in the survey data, t_1 is defined as the survey participants filled out at baseline, whereas t_2 is defined as the survey they filled out at the 12-month follow-up. With the behavioral log data, t_1 is defined as participants' web-based engagement during the first 6 months (baseline to month 6), and t_2 is defined as their engagement during the last 6 months (from month 6 to month 12).

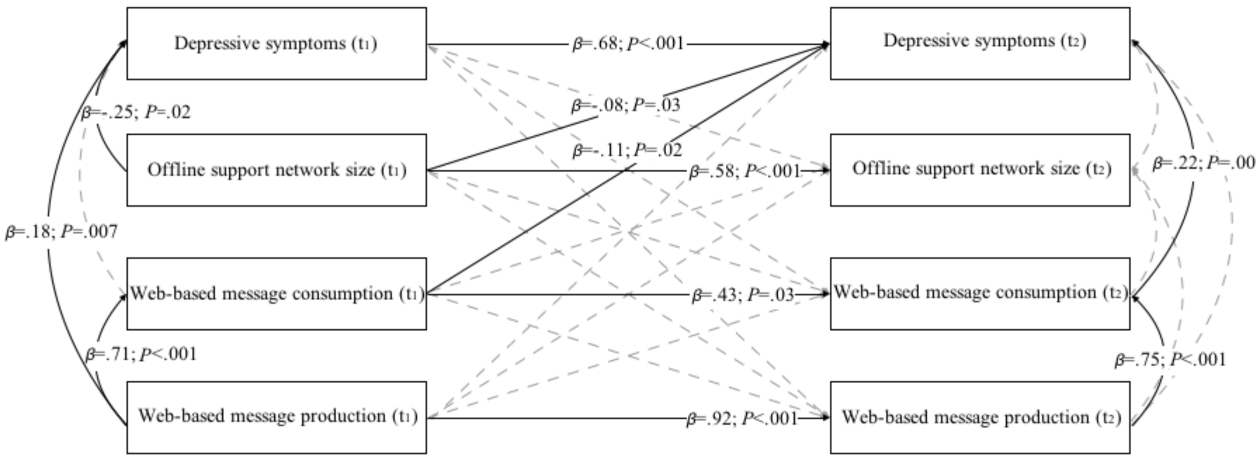
Web-based message consumption was operationalized as the number of messages participants viewed in the peer-to-peer discussion forums during the first 6 months of participation (t1: mean 1063.87, SD 3478.41) and the following 6 months (t2: mean 700.54, SD 2423.15).

Web-based message production was operationalized as the number of messages participants created in the discussion forum for the first 6 months (t1: mean 23.80, SD 78.45) and the following 6 months (t2: mean 28.11, SD 106.37).

Data Analysis

To test the hypotheses, depressive symptoms, offline support network, web-based message consumption, and web-based message production measured at t_1 were allowed to simultaneously predict these same constructs at t_2 in a cross-lagged panel design. The 4 exogenous variables at t_1 and the 4 endogenous variables at t_2 were specified to correlate. Most relevant for evaluating our hypotheses are the cross-lagged paths between the constructs assessed at both t_1 and t_2 (Figure 1). Models were tested using a full information maximum likelihood estimation, in which missing data were imputed [64]. Given that web-based message consumption and production data were highly skewed, we used maximum likelihood estimation with robust standard errors to address the nonnormality problem.

Figure 1. Cross-lagged path model of observed long-term effects on depressive symptoms, offline support network size, web-based message consumption, and web-based message production over 1 year. Nonsignificant paths are represented as dashed lines with muted colors without their coefficients.



Fully cross-lagged models require the saturated estimation of all possible path combinations, but this would produce just-identified models containing 0 degrees of freedom. This makes it impossible to obtain model fit indices [65]. To obtain optimal model fit indices and achieve parsimony, it is recommended to undertake post hoc modification, whereby nonsignificant and theoretically uninformative paths are removed [56]. Thus, we removed 2 paths showing insignificant relationships between exogenous variables (ie, web-based message consumption at t_1 and offline support network at t_1 and web-based message production at t_1 and offline support network at t_1).

The resulting model showed an excellent fit to the data (χ^2_{27}=273.9 root mean square error of approximation=0.04; standardized root mean square residual=0.01; comparative fit index=0.99; and Tucker-Lewis Index=0.96). The explained variance (R^2) for depressive symptoms, offline support network, web-based message consumption, and web-based message production at t_2 were 57.6%, 34.1%, 68.8%, and 81.3%, respectively.

Results

Descriptive Statistics

Descriptive characteristics of participants at baseline are presented in Table 1. Means, SDs, and person correlation coefficients between all variables that were subsequently included in the cross-lagged models are reported in Table 2. Participants' reports of depressive symptoms, offline support networks, web-based message production, and web-based message consumption did not change over time.

Table 1. Descriptive characteristics of participants at baseline (N=197).

Characteristics	Values
Age, years, mean (SD)	76.26 (7.38)
Gender, n (%)	
Male	52 (26.4)
Female	145 (73.6)
Race or ethnicity, n (%)[a]	
White	176 (89.3)
Black	19 (9.6)
Other	8 (4.1)
Education, n (%)	
Less than high school	0 (0.0)
Some high school or diploma	74 (37.6)
Some college or post-high school	68 (34.5)
4-year degree or above	55 (27.9)
Living arrangement, n (%)[a]	
Living alone	121 (61.4)
Spouse or partner	61 (31.0)
Son or daughter	15 (7.6)
Other family or friends	3 (1.5)
Paid caregiver	1 (0.5)
No response	1 (0.5)
Comfort with technology, mean (SD)[b]	
Smartphone or tablet	1.4 (1.8)
Desktop computer	3.2 (1.8)
Email	2.8 (2.1)
Facebook	1.7 (2.0)

[a] Group totals may exceed 100% because participants could report more than one race or ethnicity and living arrangement.

[b] Comfort with technology was measured with a 6-point scale (0=*never used* to 5=*very comfortable*).

Table 2. Means, SDs, and Pearson correlation coefficients for key variables at both wavesa.

Variables	Mean (SD)	1. Depressive symptoms (t_1)	2. Depressive symptoms (t_2)	3. Offline support network size (t_1)	4. Offline support network size (t_2)	5. Web-based message consumption (t_1)	6. Web-based message consumption (t_2)	7. Web-based message production (t_1)	8. Web-based message production (t_2)
1. Depressive symptoms (t_1)	4.42 (4.65)	1	__[b]	—	—	—	—	—	—
2. Depressive symptoms (t_2)	3.91 (4.29)	.76[c]	1	—	—	—	—	—	—
3. Offline support network size (t_1)	4.95 (4.80)	−.25[c]	−.28[c]	1	—	—	—	—	—
4. Offline support network size (t_2)	6.21 (8.32)	−.17[d]	−.22[e]	.61[c]	1	—	—	—	—
5. Web-based message consumption (t_1)	1063.87 (3478.41)	.03	.04	−.02	−.00	1	—	—	—
6. Web-based message consumption (t_2)	700.54 (2423.15)	.18[e]	.23[e]	−.03	−.00	.69[c]	1	—	—
7. Web-based message production (t_1)	23.80 (78.45)	.09	.10	.03	−.01	.71[c]	.70[c]	1	—
8. Web-based message production (t_2)	28.11 (106.37)	.10	.10	.04	−.01	.62[c]	.76[c]	.90[c]	1

[a] Although all variables were denoted as t_1 and t_2, the specific time points varied—for depressive symptoms and offline support network size, t_1 and t_2 indicate baseline and follow-up 12 months later, respectively. For web-based message consumption and production, t_1 and t_2 indicate the first 6 months (baseline to month 6) and the last 6 months of the study (from month 6 to month 12), respectively.

[b] The correlation coefficient was not shown as it was shown in the asymmetrically diagonal position of the table.

[c] Correlations significant at the 0.001 level.

[d] Correlations significant at the 0.05 level.

[e] Correlations significant at the 0.01 level.

Hypothesis Testing

First, the cross-lagged relationship between older adults' offline social connectedness at baseline (t_1) and their depressive symptoms a year later (t_2) was examined to replicate previous findings. Results confirmed that larger offline support networks at baseline significantly reduced older adults' depressive symptoms a year later (β=−.08; *P*=.03).

A cross-sectional effect also emerged at baseline, with offline support networks negatively associated with depressive symptoms (β=−.25; *P*=.02), but no such correlation emerged a year later (β=−.06; *P*=.22). Substantial autocorrelations were found: depressive symptoms at baseline were significantly related to depressive symptoms a year later (β=.68; *P*<.001), and offline support network size at baseline was significantly associated with offline support network size a year later (β=.58; *P*<.001).

To test hypothesis 1, the cross-lagged relationship between older adults' web-based message consumption and production for the first 6 months (t_1) and their depressive symptoms a year after the study began (t_2) was examined. A statistically significant negative lagged effect of web-based message consumption on depressive symptoms emerged (β=−.11; *P*=.02), but web-based message production had no lagged effect on depressive symptoms (β=.12; *P*=.34; Figure 1). This means that the more messages older adults read on the web-based forum during the first 6 months, the less depressed they felt a year after the study began, above and beyond the effect of offline support networks at baseline. However, the number of messages they wrote on the web-based forum during the first 6 months (t_1) did not affect their depressive symptoms at the 1-year check marker (t_2). Hypothesis 1 was partially supported.

Some cross-sectional effects between web-based connectedness and depressive symptoms emerged: greater depressive symptoms at baseline were associated with increased message production during the first 6 months of the study (β=.18; *P*=.007), but no such correlation emerged during the next 6 months (β=−.18; *P*=.14). Although there was no relationship between reading messages on the discussion board during the first 6 months and depressive symptoms at baseline (β=−.10; *P*=.11), participants who read more messages during the next 6 months experienced more depressive symptoms a year after the study began (β=.22; *P*=.007).

As expected, substantial autocorrelations were also found: web-based message consumption during the first 6 months of the study was significantly related to consumption during the next 6 months (β=.43; P=.03), and web-based message production during the first 6 months was significantly associated with production during the next 6 months (β=.92; P<.001). Web-based message consumption was correlated with production at both initial assessment (β=.71; P<.001) and 1 year later (β=.75; P<.001).

To address research question 1, the standardized coefficients for the significant lagged paths (ie, the effects of web-based message consumption during the first 6 months of the study [t_1] on depressive symptoms a year after the study began [t_2] and the effects of offline network size at baseline [t_1] on depressive symptoms a year later [t_2]) were compared [66]. Web-based message consumption had a considerably larger effect (β=−.11; P=.02) in reducing depressive symptoms 1 year after participation in the support website, compared with offline support network size (β=−.08; P=.03; Figure 1). To confirm that these 2 effects were in fact statistically different, the joint hypothesis test (*F* test) was performed. The results indicated that the difference between the 2 effects was significant ($F_{1,141}$=55.242; P<.001).

Hypothesis 2 dealt with the issue of chain of causality, specifically whether depressive symptoms predict (in addition to being affected by) web-based connectedness. The results show no lagged effect of depressive symptoms on either web-based message consumption (β=.12; P=.36) or production (β=.02; P=.43), meaning that depressive symptoms at baseline did not play a role in older adults' web-based connectedness during the next 6 months. Hypothesis 2 was not supported.

The data also allowed us to make inferences about the nature of the relationship between depressive symptoms and *offline* connectedness. The results show no cross-lagged effect of depressive symptoms on offline support networks (β=−.04; P=.43), meaning that baseline depressive symptoms did not lead older adults to lose their offline support networks over time.

Discussion

Principal Findings

The goal of this study was to understand the protective value of web-based social connectedness (ie, the extent to which older adults engaged with each other on a support website) against older adults' depressiveness. We captured behavioral trace data (ie, the number of messages participants wrote and read on the support website) and participants' self-reported depressiveness over the course of a year. The findings can be summarized as follows. Replicating prior research, the availability of offline support networks decreased older adults' depressive symptoms a year later. Most importantly, aspects of web-based connectedness *further* reduced older adults' depressive symptoms over time. Specifically, web-based message consumption, or the extent to which older adults read peers' messages on the support website, reduced their depressive symptoms above and beyond the reduction generated by offline social connectedness. In fact, web-based message consumption had a larger effect than the availability of offline support networks in reducing older adults' depressiveness. However, contrary to predictions, web-based message production, or the extent to which older adults wrote messages on the online discussion forum, did not affect their depressive symptoms a year later. Finally, initial depressive symptoms did not affect web-based message consumption or production, meaning that individuals who were more depressed initially did not seek out opportunities for social connection on the support website.

Theoretical Implications

These findings contribute to the literature on several fronts. First, we advance research on older adults' social connectedness and depressive symptoms, which, to date, has focused on traditional connections that originate in face-to-face settings (eg, family members and friends from the community). By shifting the lens toward *web-based* opportunities for connection, we found that low-effort participation on a support website, for instance simply reading messages generated by others experiencing similar life circumstances, afforded older adults significant protection against depressive symptoms. Our data do not address the mechanism behind this effect. However, likely candidates are suggested by research on the benefits of online social support. The key feature of support websites is that they operate via *homophily*, meaning that they connect individuals experiencing similar predicaments [18,34,47]. Although traditional face-to-face support networks, such as family members, typically wish to help, they often fail because they do not have lived experiences that would enable them to provide support seekers with a sense of being understood and validated in their struggles. For instance, research shows that younger individuals have a hard time understanding the challenges associated with aging [66]. Conversely, homophily-based networks by definition revolve around shared experiences. Simply witnessing others' struggles and noting parallels with their own likely make older adults feel less alone [44] and part of a meaningful community [48]. In turn, feelings of inclusion and validation may reduce older adults' depressive symptoms. This explanation is bolstered by the finding that web-based message consumption had even stronger effects in terms of reducing older adults' depressiveness than the availability of offline support networks. Research finds that homophily increases mutual empathy and commitment to the group [34], explaining why older adults might feel a powerful sense of belonging with and validation from the virtual strangers who share similar experiences on the web. In addition, these virtual strangers became, over time, a daily presence in participants' lives, with data showing that the majority of participants logged into the site on a daily basis. Even in the absence of actual interaction, this daily surveillance likely produced strong feelings of rapport. By contrast, older adults report significantly less frequent interactions with the members of their offline support networks, such as family and offline friends [67].

A related explanation for the benefits of web-based message consumption is that older adults in a support website could accrue informational support by taking the role of a passive user [47]. Informational support refers to advice, insights, tips, and other information that can help individuals manage difficult situations [18,68]. For example, older adults may learn about helpful medications, fitness activities, or community resources that can assist them. In turn, this informational support may help older adults adopt positive coping strategies, which could ultimately ameliorate their depressive symptoms.

Contrary to expectations, message production, or the extent to which older adults wrote messages on the online forum, did not buffer depressive symptoms. It is possible that older adults did not benefit from message production simply because they did not write much, preferring instead the less effortful activity of message consumption. Of the 197 participants, approximately 59 (29.9%) did not write any messages during the first 6 months of the study. The vast majority of messages (3774/4186, 90%) during this study period were written by only 10% (20/197) of the participants. It is likely that participants who wrote messages did, in fact, benefit from this activity. However, the small sample size of message writers in our dataset prevents us from statistically testing this possibility. Another possibility is that the therapeutic effect of writing does not last in the long run. Despite studies showing that writing helps immediately and as long as 2 weeks after writing [41,42], these benefits might dissipate over the course of a year.

Finally, our cross-lagged design was able to ask the reciprocal question of whether web-based social support was used as a resource by individuals who experienced greater depressive symptoms. The central tenet of the social compensation hypothesis [53] is that individuals use web-based communication to compensate for difficulties or missed opportunities in face-to-face settings. Extant research suggests that those manifesting depressive symptomology have a tendency to withdraw from offline interactions, either because they find them depleting or because they are shunned by face-to-face networks [20-22]; thus, web-based communication might serve as a resource for compensating for this social isolation. Neither of these predictions was borne out by these data. Depressive symptoms did not cause older adults to lose their offline support networks, nor did depressive symptoms increase the extent to which they engaged in web-based communication on the support website. Therefore, our finding that older adults with depressive symptoms did not gravitate toward web-based connectedness to satisfy social needs could be due to the fact that they did not have a shortfall to compensate for. In summary, this pattern of results suggests that depressive symptoms in older adults are best positioned as an outcome, rather than an antecedent, of web-based connectedness.

Practical Implications for Online Social Support Group Design and Use

Online support groups demonstrated effectiveness in reducing older adults' depressive symptoms in this study, indicating that this population should be encouraged to use them. Relatedly, health care systems might consider incorporating online support groups into programs that support the well-being of older adults, particularly those who are not able to attend face-to-face sessions. When it comes to the design of these support websites, our pattern of results, where message consumption (but not production) reduced older adults' depressiveness over time, suggest that the presence of lurkers should be encouraged. Although active posting on these websites is certainly necessary, the lack of posting might not be as big of a problem as people think. Even lurking can have psychological benefits for older adults dealing with depressive symptoms.

Providing coaching to older adults on how to write effective web-based messages may encourage participants' involvement in message production, from which they may obtain therapeutic benefits [41]. As suggested above, the nonsignificant effect of message production on depressive symptoms could be a consequence of most participants not writing messages. Coaching older adults may increase the number of messages they post in online support groups, allowing participants to be exposed to more of these beneficial messages.

These findings are even more important because of the COVID-19 virus. If older adults are isolated from their children and grandchildren for extended periods, this may have a negative impact on them. It is comforting to find that physical presence is not necessarily the only means by which support can be offered. The longer we have to maintain some forms of social distancing, the more important web-based vehicles for support are likely to become.

Limitations and Future Directions

Several limitations should be addressed by future research. We operationalized social connectedness as the *quantity* of web-based (ie, number of messages read and posted) and offline (ie, number of individuals in support networks) engagement. Future research should complement our findings by investigating the content of web-based messages being read and written as well as the type of interactions older adults have with their offline networks. This will provide evidence of the mechanism through which social connectedness ameliorates depressive symptoms in older adults.

In this study, we used the label *offline* to denote relationships that were likely initiated offline (eg, with family members, work friends, neighbors, or community members), not relationships that are necessarily managed offline. It is possible that older adults' communication with these networks occurs via mediated means, such as the telephone. Future research on how web-based connectedness can mitigate depressiveness among older adults should investigate mediated interactions with family members and friends made offline.

As participants were recruited on a rolling basis, they reported their depressive symptoms at different times. Given that there is evidence for seasonal fluctuations in mood or rates of depression [69], seasonality of depression may act as a potential confounder.

Finally, most participants in our sample were not clinically depressed. Only a fraction of our sample scored higher than the cutoff for clinical depression on the PHQ-8 questionnaire. However, the PHQ also captures individuals' tendency to experience depressive symptoms that fall below the severity threshold for clinical depression, which was the case for our participants. Hence, we used the terminology *depressive symptoms* rather than *depression* throughout this manuscript. Future work should identify older adults with clinical depression to ascertain if the present findings hold for this more vulnerable group.

Conclusions

Our research provides new insights into the protective power of web-based social connectedness against depressive symptoms among older adults by showing that reading others' messages on a support website led to an amelioration of older adults' depressive symptoms after 1 year of participation. This indicates that support websites might serve as a useful venue where older adults build an alternative form of social connection, which might be tremendously helpful for their psychological well-being.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary sample messages.
[DOCX File , 118 KB - jmir_v23i1e21275_app1.docx]

Multimedia Appendix 2
Supplementary descriptions about cross-lagged models.
[DOCX File , 61 KB - jmir_v23i1e21275_app2.docx]

References

1. United States Census Bureau. An Aging Nation: Projected Number of Children and Older Adults. 2018 Mar 13. URL: https://www.census.gov/library/visualizations/2018/comm/historic-first.html [accessed 2020-06-09]
2. Wan H, Goodkind D, Kowal P. An Ageing World: 2015. United States Census Bureau. 2016. URL: https://www.census.gov/content/dam/Census/library/publications/2016/demo/p95-16-1.pdf [accessed 2020-06-09]
3. Centers for Disease Control and Prevention. The State of Mental Health and Aging in America. URL: https://www.cdc.gov/aging/pdf/mental_health.pdf [accessed 2020-06-09]
4. Chapman DP, Perry GS. Depression as a major component of public health for older adults. Prev Chronic Dis 2008 Jan;5(1):A22 [FREE Full text] [Medline: 18082011]
5. An S, Jung H, Lee S. Moderating effects of community social capital on depression in later years of life: a latent interaction model. Clin Gerontol 2019;42(1):70-79. [doi: 10.1080/07317115.2018.1516263] [Medline: 30235072]
6. Mellor D, Firth L, Moore K. Can the internet improve the well-being of the elderly? Ageing Int 2008 Feb 9;32(1):25-42. [doi: 10.1007/s12126-008-9006-3]
7. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
8. Hegeman JM, van Fenema EM, Comijs HC, Kok RM, van der Mast RC, de Waal MW. Effect of chronic somatic diseases on the course of late-life depression. Int J Geriatr Psychiatry 2017 Jul;32(7):779-787. [doi: 10.1002/gps.4523] [Medline: 27273023]
9. Ho CS, Jin A, Nyunt MS, Feng L, Ng TP. Mortality rates in major and subthreshold depression: 10-year follow-up of a Singaporean population cohort of older adults. Postgrad Med 2016 Sep;128(7):642-647. [doi: 10.1080/00325481.2016.1221319] [Medline: 27500920]
10. Sivertsen H, Bjørkløf GH, Engedal K, Selbæk G, Helvik A. Depression and quality of life in older persons: a review. Dement Geriatr Cogn Disord 2015;40(5-6):311-339 [FREE Full text] [doi: 10.1159/000437299] [Medline: 26360014]
11. Sinclair TJ, Grieve R. Facebook as a source of social connectedness in older adults. Comput Human Behav 2017 Jan;66:363-369. [doi: 10.1016/j.chb.2016.10.003] [Medline: 30872122]
12. Goswami S, Köbler F, Leimeister J, Krcmar H. Using online social networking to enhance social connectedness and social support for the elderly. 2010 Presented at: International Conference on Information Systems; 2010; St Louis, MO p. 12-15 URL: http://aisel.aisnet.org/icis2010_submissions/109
13. Oxman T, Berkman L, Kasl S, Freeman JD, Barrett J. Social support and depressive symptoms in the elderly. Am J Epidemiol 1992 Feb 15;135(4):356-368. [doi: 10.1093/oxfordjournals.aje.a116297] [Medline: 1550090]
14. Teo AR, Choi H, Andrea SB, Valenstein M, Newsom JT, Dobscha SK, et al. Does mode of contact with different types of social relationships predict depression in older adults? Evidence from a nationally representative survey. J Am Geriatr Soc 2015 Oct;63(10):2014-2022 [FREE Full text] [doi: 10.1111/jgs.13667] [Medline: 26437566]
15. Domènech-Abella J, Lara E, Rubio-Valera M, Olaya B, Moneta MV, Rico-Uribe LA, et al. Loneliness and depression in the elderly: the role of social network. Soc Psychiatry Psychiatr Epidemiol 2017 Apr;52(4):381-390. [doi: 10.1007/s00127-017-1339-3] [Medline: 28154893]

16. Chiao C, Weng L, Botticello AL. Social participation reduces depressive symptoms among older adults: an 18-year longitudinal analysis in Taiwan. BMC Public Health 2011 May 10;11:292 [FREE Full text] [doi: 10.1186/1471-2458-11-292] [Medline: 21569285]
17. Anderson M, Perrin A. Tech adoption climbs among older adults. Pew Research Center. URL: http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2019-01-05]
18. Wright K. Communication in health-related online social support groups/communities: a review of research on predictors of participation, applications of social support theory, and health outcomes. Rev of Commun Res 2016;4:65-87. [doi: 10.12840/issn.2255-4165.2016.04.01.010]
19. Cruwys T, Dingle GA, Haslam C, Haslam SA, Jetten J, Morton TA. Social group memberships protect against future depression, alleviate depression symptoms and prevent depression relapse. Soc Sci Med 2013 Dec;98:179-186. [doi: 10.1016/j.socscimed.2013.09.013] [Medline: 24331897]
20. Coyne JC. Depression and the response of others. J Abnorm Psychol 1976 Apr;85(2):186-193. [doi: 10.1037//0021-843x.85.2.186] [Medline: 1254779]
21. Slavin LA, Rainer KL. Gender differences in emotional support and depressive symptoms among adolescents: a prospective analysis. Am J Community Psychol 1990 Jun;18(3):407-421. [doi: 10.1007/BF00938115] [Medline: 2264557]
22. Joiner TE, Metalsky GI. A prospective test of an integrative interpersonal theory of depression: a naturalistic study of college roommates. J Pers Soc Psychol 1995 Oct;69(4):778-788. [doi: 10.1037//0022-3514.69.4.778] [Medline: 7473031]
23. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. Psychol Aging 2010 Jun;25(2):453-463 [FREE Full text] [doi: 10.1037/a0017216] [Medline: 20545429]
24. Robitaille A, Orpana H, McIntosh CN. Reciprocal relationship between social support and psychological distress among a national sample of older adults: an autoregressive cross-lagged model. Can J Aging 2012 Mar;31(1):13-24. [doi: 10.1017/S0714980811000560] [Medline: 22321442]
25. Hawkley LC, Hughes ME, Waite LJ, Masi CM, Thisted RA, Cacioppo JT. From social structural factors to perceptions of relationship quality and loneliness: the Chicago health, aging, and social relations study. J Gerontol B Psychol Sci Soc Sci 2008 Nov;63(6):S375-S384 [FREE Full text] [doi: 10.1093/geronb/63.6.s375] [Medline: 19092047]
26. Santini ZI, Fiori KL, Feeney J, Tyrovolas S, Haro JM, Koyanagi A. Social relationships, loneliness, and mental health among older men and women in Ireland: a prospective community-based study. J Affect Disord 2016 Nov 01;204:59-69. [doi: 10.1016/j.jad.2016.06.032] [Medline: 27337705]
27. Cohen S. Social relationships and health. Am Psychol 2004;59(8):676-684. [doi: 10.1037/0003-066x.59.8.676]
28. Walther J, Boyd S. Attraction to computer-mediated social support. In: Communication Technology and Society: Audlence Adoption and Uses (New Media: Policy & Social Research Issues). Cresskill, NJ: Hampton Press; 2002:153-188.
29. Namkoong K, Shah DV, Gustafson DH. Offline social relationships and online cancer communication: effects of social and family support on online social network building. Health Commun 2017 Nov;32(11):1422-1429 [FREE Full text] [doi: 10.1080/10410236.2016.1230808] [Medline: 27824257]
30. Han JY, Shah DV, Kim E, Namkoong K, Lee S, Moon TJ, et al. Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. Health Commun 2011 Mar;26(2):185-197 [FREE Full text] [doi: 10.1080/10410236.2010.544283] [Medline: 21318917]
31. Namkoong K, Shah DV, Han JY, Kim SC, Yoo W, Fan D, et al. Expression and reception of treatment information in breast cancer support groups: how health self-efficacy moderates effects on emotional well-being. Patient Educ Couns 2010 Dec;81 Suppl:S41-S47 [FREE Full text] [doi: 10.1016/j.pec.2010.09.009] [Medline: 21044825]
32. Yoo W, Namkoong K, Choi M, Shah DV, Tsang S, Hong Y, et al. Giving and receiving emotional support online: communication competence as a moderator of psychosocial benefits for women with breast cancer. Comput Human Behav 2014 Jan;30:13-22 [FREE Full text] [doi: 10.1016/j.chb.2013.07.024] [Medline: 24058261]
33. Rains S, Young V. A meta-analysis of research on formal computer mediated support groups examining group characteristics and health outcomes. Human Commun Res 2009 Jul;35(3):309-336. [doi: 10.1111/j.1468-2958.2009.01353.x]
34. Wang Z, Walther JB, Pingree S, Hawkins RP. Health information, credibility, homophily, and influence via the internet: web sites versus discussion groups. Health Commun 2008 Jul;23(4):358-368. [doi: 10.1080/10410230802229738] [Medline: 18702000]
35. Hadjiconstantinou M, Byrne J, Bodicoat DH, Robertson N, Eborall H, Khunti K, et al. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. J Med Internet Res 2016 Oct 21;18(10):e270 [FREE Full text] [doi: 10.2196/jmir.5991] [Medline: 27769955]
36. Mo PK, Coulson NS. Online support group use and psychological health for individuals living with HIV/AIDS. Patient Educ Couns 2013 Dec;93(3):426-432. [doi: 10.1016/j.pec.2013.04.004] [Medline: 23669152]
37. Nimrod G. Seniors' online communities: a quantitative content analysis. Gerontologist 2010 Jun;50(3):382-392. [doi: 10.1093/geront/gnp141] [Medline: 19917645]
38. Pfeil U, Zaphiris P. Patterns of empathy in online communication. 2007 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 28-May 3; San Jose. [doi: 10.1145/1240624.1240763]

39. Pfeil U, Zaphiris P. Investigating social network patterns within an empathic online community for older people. Comput Human Behav 2009 Sep;25(5):1139-1155. [doi: 10.1016/j.chb.2009.05.001]
40. Campbell RS, Pennebaker JW. The secret life of pronouns: flexibility in writing style and physical health. Psychol Sci 2003 Jan;14(1):60-65. [doi: 10.1111/1467-9280.01419] [Medline: 12564755]
41. Pennebaker JW. Writing about emotional experiences as a therapeutic process. Psychol Sci 1997 May 01;8(3):162-166. [doi: 10.1111/j.1467-9280.1997.tb00403.x]
42. Pennebaker JW, Seagal JD. Forming a story: the health benefits of narrative. J Clin Psychol 1999 Oct;55(10):1243-1254. [doi: 10.1002/(SICI)1097-4679(199910)55:10<1243::AID-JCLP6>3.0.CO;2-N] [Medline: 11045774]
43. Kim E, Han JY, Moon TJ, Shaw B, Shah DV, McTavish FM, et al. The process and effect of supportive message expression and reception in online breast cancer support groups. Psychooncology 2012 May;21(5):531-540 [FREE Full text] [doi: 10.1002/pon.1942] [Medline: 21416553]
44. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
45. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. Soc Issues Policy Rev 2017 Jan 13;11(1):274-302. [doi: 10.1111/sipr.12033]
46. Toma C. Effects of Facebook self-presentation on emotional well-being. In: The Routledge Handbook of Media Use and Well-Being. New York and London: Routledge; 2016:170-182.
47. Wright K. Computer-mediated social support, older adults, and coping. J Commun 2000 Sep;50(3):100-118. [doi: 10.1111/j.1460-2466.2000.tb02855.x]
48. Nonnecke B, Preece J. Silent Participants: Getting to Know Lurkers Better. London: Springer; 2003:110-132.
49. Grieve R, Indian M, Witteveen K, Anne Tolan G, Marrington J. Face-to-face or Facebook: can social connectedness be derived online? Comput Human Behav 2013 May;29(3):604-609. [doi: 10.1016/j.chb.2012.11.017]
50. Cockshaw W, Shochet I. The link between belongingness and depressive symptoms: an exploration in the workplace interpersonal context. Aust Psychol 2010 Dec 31;45(4):283-289. [doi: 10.1080/00050061003752418]
51. Shochet IM, Homel R, Cockshaw WD, Montgomery DT. How do school connectedness and attachment to parents interrelate in predicting adolescent depressive symptoms? J Clin Child Adolesc Psychol 2008 Jul;37(3):676-681. [doi: 10.1080/15374410802148053] [Medline: 18645757]
52. Toma CL, Hancock JT. Self-affirmation underlies Facebook use. Pers Soc Psychol Bull 2013 Jan 28;39(3):321-331. [doi: 10.1177/0146167212474694]
53. Desjarlais M, Willoughby T. A longitudinal study of the relation between adolescent boys and girls' computer use with friends and friendship quality: Support for the social compensation or the rich-get-richer hypothesis? Comput Human Behav 2010 Sep;26(5):896-905. [doi: 10.1016/j.chb.2010.02.004]
54. Ruppel EK, McKinley CJ. Social support and social anxiety in use and perceptions of online mental health resources: exploring social compensation and enhancement. Cyberpsychol Behav Soc Netw 2015 Aug;18(8):462-467. [doi: 10.1089/cyber.2014.0652] [Medline: 26252931]
55. Gustafson DH, McTavish F, Gustafson DH, Mahoney JE, Johnson RA, Lee JD, et al. The effect of an information and communication technology (ICT) on older adults' quality of life: study protocol for a randomized control trial. Trials 2015;16:191 [FREE Full text] [doi: 10.1186/s13063-015-0713-2] [Medline: 25909465]
56. Zapf D, Dormann C, Frese M. Longitudinal studies in organizational stress research: a review of the literature with reference to methodological issues. J Occup Health Psychol 1996 Apr;1(2):145-169. [doi: 10.1037//1076-8998.1.2.145] [Medline: 9547043]
57. Clegg C, Jackson P, Wall T. The potential of cross-lagged correlation analysis in field research. J Occup Health Psychol 1997 Sep;50(3):177-196. [doi: 10.1111/j.2044-8325.1977.tb00374.x]
58. Newsom J. Longitudinal Structural Equation Modeling: A Comprehensive Introduction. New York, NY: Routledge; 2015.
59. Shadish W, Cook T, Campbell D. Experimental and Quasi-experimental Designs for Generalized Causal Inference. Boston, MD: Houghton Mifflin; 2002:0395615569.
60. Kenny DA. Cross-lagged panel correlation: a test for spuriousness. Psychol Bull 1975;82(6):887-903. [doi: 10.1037/0033-2909.82.6.887]
61. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
62. Kline R. Principles and Practices of Structural Equation Modeling. 2nd edition. New York, NY: Guilford; 2005:1593850751.
63. Wilcox R. Trimming and winsorization. In: Encyclopedia of Biostatistics. Hoboken, New Jersey: John Wiley & Sons; Jul 2005:8.
64. Tabachnick B, Fidell L. Using Multivariate Statistics. 5th edition. Boston, MA: Pearson/Allyn & Bacon; 2007:9780134790541.
65. Ullman JB. Structural equation modeling: reviewing the basics and moving forward. J Pers Assess 2006 Aug;87(1):35-50. [doi: 10.1207/s15327752jpa8701_03] [Medline: 16856785]

66. Kite ME, Stockdale GD, Whitley BE, Johnson BT. Attitudes toward younger and older adults: an updated meta-analytic review. J Social Issues 2005 Jun;61(2):241-266. [doi: 10.1111/j.1540-4560.2005.00404.x]
67. Cornwell B. Age trends in daily social contact patterns. Res Aging 2011 Jul 25;33(5):598-631. [doi: 10.1177/0164027511409442]
68. Ballantine PW, Stephenson RJ. Help me, I'm fat! Social support in online weight loss networks. J Consumer Behav 2011 Dec 23;10(6):332-337. [doi: 10.1002/cb.374]
69. Pjrek E, Baldinger-Melich P, Spies M, Papageorgiou K, Kasper S, Winkler D. Epidemiology and socioeconomic impact of seasonal affective disorder in Austria. Eur Psychiatry 2016 Feb;32:28-33. [doi: 10.1016/j.eurpsy.2015.11.001] [Medline: 26802981]

Abbreviations

PHQ: Patient Health Questionnaire

Edited by G Eysenbach; submitted 09.06.20; peer-reviewed by A Teo, S Tyrovolas, T McCall; comments to author 09.10.20; revised version received 29.10.20; accepted 21.11.20; published 13.01.21.

Please cite as:
Hwang J, Toma CL, Chen J, Shah DV, Gustafson D, Mares ML
Effects of Web-Based Social Connectedness on Older Adults' Depressive Symptoms: A Two-Wave Cross-Lagged Panel Study
J Med Internet Res 2021;23(1):e21275
URL: http://www.jmir.org/2021/1/e21275/
doi:10.2196/21275
PMID:33439143

Original Paper

Digital Intervention With Lifestyle Coach Support to Target Dietary and Physical Activity Behaviors of Adults With Nonalcoholic Fatty Liver Disease: Systematic Development Process of VITALISE Using Intervention Mapping

Kate Hallsworth[1,2,3], BSc, PhD; Stuart McPherson[1,2,3], BSc, MBChB, MD, FRCP; Quentin M Anstee[1,2,3], BSc, MBBS, PhD, MRCP, FRCP; Darren Flynn[4], BSc, MSc, PhD, CPsychol; Laura Haigh[1,2,3], BSc, PGDip, MPH; Leah Avery[3,4], BSc, MSc, CPsychol, PhD

[1]Liver Unit, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom
[2]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne, United Kingdom
[3]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom
[4]Centre for Rehabilitation, School of Health & Life Sciences, Teesside University, Middlesbrough, United Kingdom

Corresponding Author:
Leah Avery, BSc, MSc, CPsychol, PhD
Centre for Rehabilitation
School of Health & Life Sciences
Teesside University
Centuria Building
Middlesbrough
United Kingdom
Phone: 44 1642384130 ext 4130
Email: leah.avery@tees.ac.uk

Abstract

Background: Nonalcoholic fatty liver disease (NAFLD) is linked to excessive calorie consumption, physical inactivity, and being overweight. Patients with NAFLD can halt or decelerate progression and potentially reverse their condition by changing their lifestyle behavior. International guidelines recommend the use of lifestyle interventions; however, there remains a discordance between published guidelines and clinical practice. This is primarily due to a lack of NAFLD-specific interventions to support weight loss and improve liver function.

Objective: This study aims to use intervention mapping to systematically develop a digital intervention to support patients with NAFLD to initiate and maintain changes in their dietary and physical activity behavior to promote weight loss.

Methods: Intervention mapping consisted of 6 steps: step 1 involved a needs assessment with primary and secondary health care professionals (HCPs) and patients with NAFLD; step 2 involved identification of the social cognitive determinants of change and behavioral outcomes of the intervention; step 3 involved linking social cognitive determinants of behavioral outcomes with behavior change techniques to effectively target dietary and physical activity behavior; step 4 involved the development of a prototype digital intervention that integrated the strategies from step 3, and the information content was identified as important for improving knowledge and skills from steps 1 and 2; step 5 involved the development of an implementation plan with a digital provider of lifestyle behavior change programs to patients with NAFLD using their delivery platform and lifestyle coaches; and step 6 involved piloting the digital intervention with patients to obtain data on access, usability, and content.

Results: A digital intervention was developed, consisting of 8 modules; self-regulatory tools; and provision of telephone support by trained lifestyle coaches to help facilitate behavioral intention, enactment, and maintenance. A commercial provider of digital lifestyle behavior change programs enrolled 16 patients with NAFLD to the prototype intervention for 12 consecutive weeks. A total of 11 of the 16 participants successfully accessed the intervention and continued to engage with the content following initial log-in (on average 4 times over the piloting period). The most frequently accessed modules were *welcome to the program*, *understanding NAFLD*, and *food and NAFLD*. Goal setting and self-monitoring tools were accessed on 22 occasions (4 times per tool on average). A total of 3 out of 11 participants requested access to a lifestyle coach.

Conclusions: Intervention mapping provided a systematic methodological framework to guide a theory- and evidence-informed co-design intervention development process for patients and HCPs. The digital intervention with remote support by a lifestyle coach was acceptable to patients with NAFLD and feasible to deliver. Issues with initial access, optimization of information content, and promoting the value of remote lifestyle coach support require further development ahead of future research to establish intervention effectiveness.

(J Med Internet Res 2021;23(1):e20491) doi:10.2196/20491

KEYWORDS

nonalcoholic fatty liver disease; internet-based intervention; lifestyle; diet; physical activity; weight loss

Introduction

Background

Nonalcoholic fatty liver disease (NAFLD) affects up to 33% of adults in western countries and is the most common liver condition worldwide [1]. Approximately 40% of patients with NAFLD will develop progressive liver fibrosis, and ultimately, 5% to 11% of patients with NAFLD will develop end-stage liver disease [2,3]. Weight loss, achieved through changes in dietary and physical activity behaviors, is the recommended treatment for NAFLD, which can reduce liver fat, inflammation, and fibrosis [4,5]. Evidence-based clinical guidelines for the management of NAFLD state the importance of lifestyle behavior change in patients with NAFLD, regardless of disease severity [6-8]. However, patients with NAFLD typically have their condition monitored rather than actively managed with minimal support to change their lifestyle behaviors [9]. A potential explanation for this evidence-practice gap is the lack of NAFLD-specific, theory- and evidence-informed interventions for use in routine clinical practice.

Intervention Development

Intervention mapping provides a robust method for systematically developing theory- and evidence-informed interventions that integrates the perspectives and needs of the target populations [10]. It consists of 6 consecutive steps to provide a transparent account of the translation of theory- and evidence-based behavior change techniques (BCTs) into intervention components, which explicitly target theoretical determinants of behavior and behavior change [10]. This process enables replication in terms of development and delivery of interventions and facilitates intervention optimization and robust evaluation.

In the context of this study, the target populations of interest are patients with NAFLD and health care professionals (HCPs) responsible for their care. Our previous research indicated that lifestyle behavior change interventions need to be capable of meeting the needs of both patients and HCPs. This increases the likelihood that HCPs will use or refer patients to the intervention, that uptake by patients is optimal, and that the intervention is capable of successfully improving outcomes for patients with NAFLD in the long term [9-11].

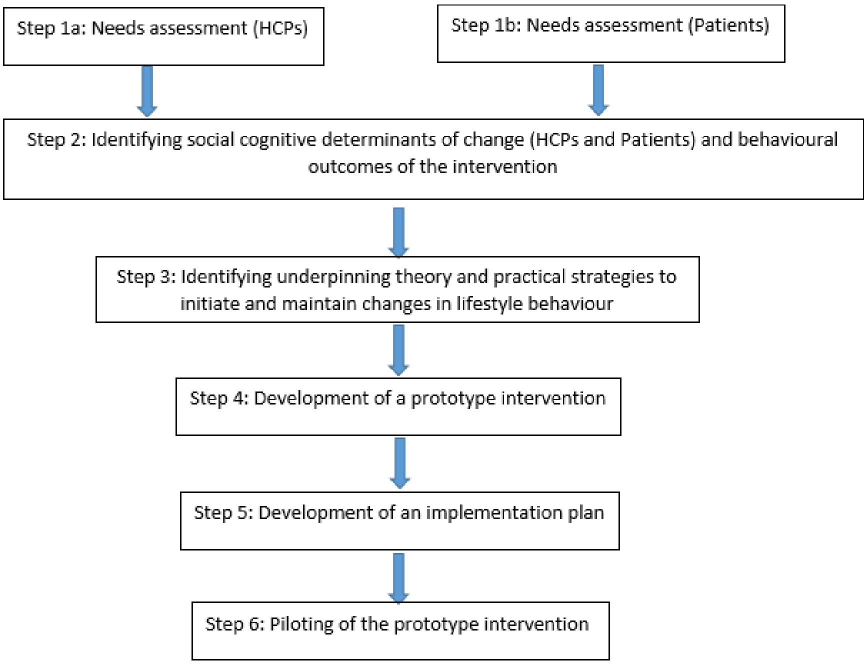
We present a detailed overview of the systematic development of an NAFLD-specific digital intervention using intervention mapping [10]—VITALISE (Intervention to Promote Lifestyle Change in Nonalcoholic Fatty Liver Disease). We describe the co-design process with patients and HCPs used to develop the form and information content of VITALISE to target dietary and physical activity behaviors to support weight loss and weight loss maintenance. This enhances the transparency and replicability of the intervention, facilitates optimization post piloting, and increases the likelihood of uptake and sustainability of the intervention over time.

Methods

Overview of Development Process

A diagrammatic summary of the co-design process that corresponds to the 6 stages of intervention mapping [10] is shown in Figure 1. Ethical approval for all stages was secured from the National Health Service (NHS) London-Riverside Research Ethics Committee (reference: 15/LO/0815).

Figure 1. Overview of the intervention development process undertaken with reference to intervention mapping. HCP: health care professional.



Step 1: Needs Assessment

Step 1 of intervention mapping involved a needs assessment of the target populations (ie, HCPs responsible for diagnosing and managing patients with NAFLD and patients diagnosed with NAFLD). We referred to national and international guidelines for the diagnosis and management of NAFLD [6-8] to inform the development of topic guides for semistructured one-to-one interviews with 21 HCPs and 12 patients. Interviews were conducted either by a senior physiotherapist with expertise in NAFLD or a health psychologist with expertise in health behavior change. Both thematically analyzed the data generated. The methods and findings of these interviews have been published previously [9].

Step 2: Defining the Objectives and Behaviors to Change (HCPs and Patients)

Step 2 involved exploring the social cognitive determinants (eg, attitudes, risk perceptions, and self-efficacy) of behavioral intention and behavior change (performance objectives) identified from the interviews with HCPs and patients during step 1. This included defining the behavioral outcomes of the intervention (change objectives, eg, dietary and physical activity behavior). Specifically, the change objectives describe how the social cognitive determinants of behavioral intention and behavior can be targeted to inform selection of practical strategies that should be used to target these determinants in step 3.

Step 3: Identifying Underpinning Theory and Practical Strategies to Initiate and Maintain Changes in Lifestyle Behavior

Step 3 involved linking social cognitive determinants of intention and behavior (eg, attitudes, beliefs, and risk perception) with the behavioral outcomes (informed by steps 1 and 2) and selecting appropriate theory- and evidence-based intervention strategies [12] to target the behaviors of interest (ie, diet and physical activity). Selection of specific BCTs was informed by the findings of our qualitative research with HCPs and patients and involved identification of specific theoretical domains using the Theoretical Domains Framework (TDF) [11-13]. Domains identified from the perspectives of HCPs (eg, professional role and identity, skills, and knowledge) and patients (eg, knowledge, skills, and self-regulation) informed the selection of BCTs during team discussions using a valid and reliable taxonomy [12]. For example, the theoretical domain *knowledge* led to the selection of the BCTs' *information about health consequences* and *information about antecedents* and the theoretical domain *professional role and identity* led to the inclusion of lifestyle coach support. Mapping the intervention strategies in this way enabled identification of what is likely to work and not work and serves as a mechanism to optimize the content of the intervention.

Step 4: Development of a Prototype Intervention

In collaboration with a design team at Changing Health Limited [14], step 4 involved the development of a prototype digital intervention that integrated the theory-based practical strategies selected during step 3 to target dietary and physical activity behavior of patients to promote weight loss. A draft paper-based overview of VITALISE was developed by the research team to effectively communicate the content to the designers, which was subsequently developed into a digital draft of the intervention.

HCPs and patients with NAFLD were further engaged separately in interactive group workshops (consisting of a demonstration of the draft digital intervention, small group work, and plenary discussion) to elicit their views and preferences on the form and information content to develop a prototype intervention and to identify any issues related to acceptability and usability.

A total of 4 workshops were conducted: two with primary HCPs (group 1 involved 4 participants and group 2 involved 5 participants), one with secondary and tertiary HCPs (involving 5 participants), and one with patients (involving 9 participants). Critically, we explored the optimal way for HCPs to describe the intervention to patients in terms of its aims, objectives, content, and delivery. In addition, we sought the views of HCPs on their support requirements to follow up patients and provide feedback. Workshops were facilitated by a senior physiotherapist and health psychologist and lasted 90 min with patients and 45 min with HCPs. Detailed notes were taken throughout the interactive group sessions.

In addition, participating HCPs were given full access to the draft digital intervention to review individually and were asked to provide feedback on the content with reference to NAFLD clinical guidelines. The outcomes of the workshops and feedback from HCPs informed refinements to the draft version of the intervention to produce an optimized version of the prototype ready for piloting.

Participants

HCPs and patients from 2 NHS hospital trusts and 11 NHS clinical commissioning groups in England were invited to take part in the study (steps 1 to 4 of the intervention mapping process). HCPs working in both hospital and community settings were purposively sampled from specialties that manage patients with NAFLD on a regular basis (including hepatology, gastroenterology, diabetology, and general practice) to ensure representation across specialties and inclusion of different professions (eg, physicians, nursing, and dietitians).

A convenience sample of patients aged ≥18 years with a confirmed diagnosis of NAFLD, regardless of disease severity, was identified by clinicians from primary, secondary or tertiary care settings. They were invited to take part in the study (steps 1 to 4 of the intervention mapping process) by a letter of invitation and asked to contact a member of the research team if they were interested in taking part.

Step 5: Development of an Implementation Plan

Changing Health Limited [14] is a private digital lifestyle behavior change organization that has a proven track record of delivering lifestyle behavior change programs, with lifestyle coach support at scale to the NHS. They agreed to host VITALISE alongside their existing programs for type 2 diabetes management and prevention. Meetings were held with Changing Health's digital development team and operations manager to plan and agree how patients would be given access to the intervention and who would provide technical support throughout the pilot testing described in step 6. In collaboration with Changing Health Limited, an implementation plan was developed to deliver VITALISE to patients with NAFLD using their delivery model (ie, digital intervention with remote support by lifestyle behavior coaches). The coaches provided by Changing Health were trained by a senior physiotherapist with specific expertise in NAFLD and a chartered health psychologist with expertise in lifestyle behavior change. The training was delivered face-to-face during one session lasting 2 hours.

Step 6: Piloting the Prototype Intervention

Step 6 of the intervention mapping process involved pilot testing VITALISE with a small volunteer sample of 16 participants recruited from LiverNorth, a national liver patient support group. Individuals were eligible to participate if they had a confirmed diagnosis of NAFLD or had been advised by an HCP that they were at risk of developing the condition due to being overweight or obese. The aim was to obtain data on access and engagement with the intervention and to obtain feedback from patients to inform optimization ahead of a larger scale evaluation. Automatically logged data on access and engagement with VITALISE were analyzed using appropriate summary statistics, with qualitative data from the feedback questionnaire subjected to iterative conceptual content analysis by the author (KH). A priori and emergent coding were used to summarize key themes for discussion with the research team, who served as a challenge forum on the integrity of the analysis.

Results

Overview of Sample

We engaged HCPs and patients in a co-design process to develop a theory-informed digital intervention—VITALISE. Details of the participants are shown in Table 1.

Table 1. Summary of demographics for health care professionals and patients who participated throughout the intervention development process.

Characteristics	Values
Health care professionals (n=21), n (%)	
Gender	
Male	10 (48)
Female	11 (52)
Specialty	
Hepatology (3×consultants; 1×specialist registrar)	4 (19)
Gastroenterology (4×consultants)	4 (19)
Diabetology (3×consultants)	3 (13)
Primary care physician	6 (29)
Dietician (1×weight management lead 1×diabetes lead)	2 (10)
Hepatology specialist nurse	1 (5)
Primary care practice nurse	1 (5)
Patients with NAFLD[a] (n=28)	
Gender, n (%)	
Male	12 (43)
Female	16 (57)
Characteristics	
Age (years), mean (SD; range)	59 (8; 42-72)
Ethnicity (White or White British), n (%)	28 (100)
Diagnosed NAFLD, n (%)	22 (79)
High risk of NAFLD due to overweight or obesity, n (%)	4 (14)
Other liver disease, n (%)	2 (7)

[a] NAFLD: nonalcoholic fatty liver disease.

Intervention Mapping Step 1: Needs Assessment With HCPs and Patients

The findings of the interviews with 21 HCPs and 12 patients have been published previously [9], and a summary of key findings is presented here.

HCPs included hepatology (3×consultants and 1×specialist registrar), gastroenterology (4 consultants), diabetology (3x consultants), primary care physicians, dieticians (1×weight management lead and 1×diabetes lead), a hepatology specialist nurse, and a primary care practice nurse. HCPs reported a lack of knowledge, skills, and available resources to effectively target lifestyle behavior changes in patients following the diagnosis of NAFLD. Several HCPs acknowledged that targeting lifestyle behavior change should be part of routine practice. However, due to time constraints and competing clinical demands, they reported advising patients to lose weight and subsequently monitored their patients' weight thereafter with no further lifestyle intervention. Nonspecialist HCPs reported a lack of knowledge about NAFLD and its management and emphasized the need for training to provide evidence-based information and advice to patients during consultations [9].

The lifestyle behavior change literature on NAFLD places an emphasis on a 7% to 10% weight loss target for patients to reduce their liver fat and inflammation sufficiently to improve liver health [4,5]. An incremental weight loss percentage increase is associated with better outcomes [15,16]. However, most patients interviewed reported being unaware of the association between weight loss and improvements in liver health. Importantly, none of the patients interviewed reported being aware that NAFLD was progressive and that lifestyle behavior changes could prevent or decelerate the progression of NAFLD to severe liver disease, and in some cases reverse the condition. When asked about their information needs at the time of diagnosis (ie, what would increase their intention to make lifestyle behavioral changes), they reported the need for information on risks associated with NAFLD and how NAFLD can be managed with dietary and physical activity behavior changes. The majority of patients reported being advised to lose weight by HCPs; however, this advice was regularly received because many patients had comorbid conditions, including obesity and diabetes. Therefore, they did not universally link weight loss advice to NAFLD, its development, or progression. None of the patients interviewed reported being supported to make behavioral changes. However, the majority emphasized that increased awareness that NAFLD could be effectively managed in this way, including guidance on what kinds of changes to make, how to make changes, and the required magnitude of changes, would strengthen their behavioral intentions for increasing physical activity and making dietary changes to reduce their energy consumption and composition of diet. The value of support from a trusted HCP was also emphasized. In this regard, information, feedback on progress, and social support were identified as important for a proportion of those interviewed; however, the evidence most strongly supported the need for information about NAFLD, risk of progression, and information about effective dietary and physical activity approaches [9]. Data were further analyzed using the TDF to identify specific targets for intervention. The findings have been published previously [11]. There is evidence to support a variety of dietary approaches to facilitate weight loss in the context of NAFLD (eg, Mediterranean diet [with a focus on increased fish, fruit, and vegetable intake] to promote weight loss and improve liver health, intermittent fasting, and calorie restriction to reduce energy intake); however, barriers exist to uptake and adherence. Therefore, to support individual preferences and to increase the likelihood of continued engagement, the intervention provided information on all 3 of these dietary approaches as well as links to trusted websites for more detailed information (eg, recipe ideas). As such, participants were encouraged to select the dietary approach that they felt would work best for them. This could involve reducing the amount of food they usually eat (eg, reduce portion sizes) or substitute food items usually consumed with alternatives (eg, substitute whole milk for semiskimmed or red meat for fish). In terms of physical activity and exercise, there were no specific recommendations regarding the type and amount of physical activity. Instead, the role of physical activity and exercise in the context of NAFLD was provided (eg, promotion of weight loss maintenance and reduction in liver fat) with additional information about different types of physical activities and exercises to promote an informed- and preference-based choice. The focus was on increasing everyday levels of activity.

Intervention Mapping Step 2: Identifying Determinants of Intention and Behavior of HCPs and Patients

A total of 8 theoretical domains (out of 14) were identified from the analyses of interviews with HCPs in step 1. These were beliefs about consequences; professional role and identity; environmental context and resources; knowledge; skills; goals; behavioral regulation; and memory, attention and decision processes [11]. Although HCPs highlighted a lack of knowledge and skills to effectively target lifestyle behavior changes in their patients, the issue of competing demands (eg, clinical goals) and lack of time emerged as salient barriers. Furthermore, a small number of HCP specialists queried whether delivery of lifestyle behavior change should be part of their role (professional role and identity), that is, the requirement to deliver interventions would potentially take them away from their core clinical role. This suggested that any intervention developed for delivery by specialists may face implementation challenges. As such, we defined the target behaviors of HCPs as promotion of lifestyle behavior change (ie, emphasizing the importance in the context of NAFLD), providing advice, referring to another professional for lifestyle behavior change support (eg, to an external provider), and providing feedback on progress in relation to the impact of behavioral changes on clinical outcomes. Although it was acknowledged that there is a need for training to achieve this (eg, knowledge of what NAFLD is, how it develops, and how lifestyle behavioral change plays a vital role in the management of the disease), this perceived burden was considered less than delivering an intervention (as opposed to referring the patient) within routine consultations. HCPs perceived their patients as resistant or reluctant to change lifestyle behaviors (beliefs about consequences); however, this was largely attributable to patients reporting a lack of awareness of the links between lifestyle changes, weight, and liver health in the context of NAFLD, which inhibited them from developing an intention to change their diet and physical activity levels.

A total of 4 theoretical domains were identified from interviews with patients in step 1 as the focus for intervention. These were knowledge, beliefs about consequences, social influences, and behavioral regulation [11]. Patients emphasized a need for information, support, and feedback on progress from a trusted HCP; therefore, the performance objectives of HCPs were extended to monitor outcomes of behavioral changes with their patients to provide them with tailored feedback (ie, feedback on the impact that lifestyle behavioral changes had made on their weight status and liver health) and support.

Therefore, the objective of the patient intervention was to communicate information to target attitudes, beliefs, and risk perceptions to increase the strength of behavioral intentions. Support with planning and monitoring dietary and physical activity changes to initiate and maintain weight loss was considered important once patients were aware of the benefits of lifestyle change. Monitoring progress in particular (behavioral regulation) was identified by patients as a process that would promote self-efficacy and further support motivation.

Intervention Mapping Step 3: Selecting Theory and BCTs

The findings of steps 1 and 2 were used to guide the selection of theory and theory-linked BCTs to structure and populate the content of the digital intervention. It became apparent upon completion of our qualitative interview study that patients were poorly informed about NAFLD, what it is, how it manifests, associated risks, and how it can be managed. Furthermore, many patients interviewed reported being advised that NAFLD was nothing to worry about. They expressed a desire to know more about NAFLD, specifically the risks of progression following their invitation to the study, that is, for many patients, it only become known by taking part in the study that NAFLD could progress to cirrhosis, for example. Only a minority of participants reported the need for support to make and sustain behavioral changes. Given the need to focus on knowledge, beliefs, and risk perception to promote intention formation, and the importance of support to achieve and maintain weight loss for a proportion of individuals interviewed, the health action process approach (HAPA) was considered an appropriate theory to underpin the development of the intervention [17].

As described in step 2, risk perception was specifically emphasized by patients as key to behavioral intention formation. Once they were aware of the potential benefits of making lifestyle behavior changes to achieve weight loss and the positive impact of this on liver health, they reported a need for support to make and sustain lifestyle behavior changes to lose weight and maintain weight loss. Table 2 provides an overview of the theoretical constructs of the HAPA, BCTs used to operationalize each construct, and a description of the mechanism underpinning each BCT.

It was important that all constructs of the HAPA were targeted with conceptually appropriate intervention components and theory-linked BCTs to fully operationalize the theory. This ensures that the intervention is replicable and can be adequately tested and increases the likelihood that the intervention will be effective for changing behaviors [18].

Table 2. Overview of the theoretical constructs, behavior change techniques used to operationalize each construct, and a description of their use within the digital intervention to promote lifestyle change in nonalcoholic fatty liver disease.

Theoretical construct	BCTs[a] to operationalize constructs [12] (corresponding taxonomy code numbers)	Description and purpose of BCT use	Source of BCT selection from TDF[b] analyses in Step 1[c]
Risk perception	• Information on antecedents (4.2) • Information about health consequences (5.1)	Communicate and challenge perceptions about NAFLD[d] risk	Domain Knowledge
Outcome expectations	• Pros and cons (9.2) • Credible source (9.1)	Present pros and cons for making lifestyle behavior changes in relation to risk and management of NAFLD. An NAFLD specialist communicates this message to increase credibility	Domains Beliefs About Consequences and Knowledge
Planning (action and coping)	• Goal setting behavior (1.1) • Social support (unspecified; 3.1) • Social support (practical; 3.2) • Action planning (1.4) • Problem solving (1.2) • Instruction on how to perform a behavior (4.1) • Behavioral substitution (8.2) • Self-monitoring of behavior (2.3) • Self-monitoring of behavior on outcomes of behavior (2.4) • Review behavior goals (1.5)	Prompt and support the development of behavioral goals and plans. Problem-solving strategies, social support, self-monitoring, and feedback promote maintenance. Review goals to enhance motivation and promote maintenance	Domains Goals, Social Influences, Knowledge, and Behavioral Regulation
Self-efficacy (task and coping, recovery)	• Social support (unspecified; 3.1) • Social support (practical; 3.2) • Self-monitoring of behavior (2.3) • Self-monitoring of behavior on outcomes of behavior (2.4) • Feedback on outcomes of behavior (2.7)	Provide mechanisms for ensuring risk is adequately understood, planning is realistic and within capabilities, and problem solving is explicitly linked to target behaviors (diet and physical activity). Feedback and self-monitoring to provide positive reinforcement	Domains Social Influences and Behavioral Regulation

[a] BCT: behavior change technique.

[b] TDF: Theoretical Domains Framework.

[c] Source of BCT selection from TDF analyses were identified from the analyses of interviews with HCPs and patients in step 1.

[d] NAFLD: nonalcoholic fatty liver disease.

Intervention Mapping Step 4: Designing the Prototype Intervention

Following discussions within the research team regarding the findings in steps 1 to 3, a consensus was reached that a digital intervention would be the optimal mode of delivery for the proposed lifestyle intervention. This would facilitate standardization of the intervention delivery across primary, secondary, and tertiary care settings. Digital delivery with remote support also offers a scalable solution that can be tailored to individual patient needs. Evidence has shown that digital interventions can effectively target social cognitive determinants of behavior [19,20], which is vitally important in the context of NAFLD management with reference to our qualitative findings.

Patients highlighted the need for personalized feedback on behavioral changes (diet and physical activity) and outcomes of behavioral changes (eg, weight and liver health); therefore, the decision to include lifestyle coaches to promote continued engagement through feedback provision (specifically feedback on behaviors) and support to plan, overcome barriers, and monitor behavior was considered important. As such, alongside digital delivery, remotely delivered (telephone) lifestyle coach support was provided to maximize engagement with the intervention and to provide the desired feedback and support required by patients. Lifestyle coach support has been demonstrated to increase engagement with digital interventions [21] and the use of self-regulation strategies to target dietary and physical activity behavior within the context of type 2 diabetes prevention and management [22], which were considered transferable in the context of NAFLD. In addition, research has reported on the benefits of remotely delivered behavior change interventions [23,24], including computer-delivered interventions for effectively targeting social cognitive determinants of behavior, including risk [19].

The content of VITALISE was informed by steps 1 to 3 and was designed to increase knowledge about NAFLD, to raise awareness about the risk of NAFLD progression, and to highlight the associations with overweight or obesity to increase motivation or intention to make behavioral changes to promote weight loss. It was clear from the needs assessment that clinical team members were unable to provide this type and level of support due to training needs and competing clinical demands. The intervention was therefore designed to complement the existing care pathway and provided a means for HCPs to refer patients for lifestyle behavior change support.

A prototype intervention was developed in collaboration with the design team at Changing Health Limited [14] and comprised 8 modules (Table 3) accessible via a home screen of the digital program (Figure 2). Participants were tunneled through the modules in a sequential order to maximize engagement with the intervention content.

Module 1 provides a message from a specialist in lifestyle management in NAFLD welcoming patients to the program and presents a summary of the aims and objectives. Module 2 presents information on NAFLD (how it is identified or diagnosed, risk of progression, understanding test results, management strategies, and the potential long-term complications of NAFLD when left unmanaged) and patient narratives reporting on success stories. Module 3 provides information on how to access a lifestyle coach, how to make the most of coaching, and how to book follow-up sessions. Module 4 is specifically about food in the context of NAFLD and describes the role of macronutrients, nutrition, and energy balance. Module 5 uses the fundamentals of module 4 to provide information on selecting an appropriate dietary approach (eg, Mediterranean diet [25] and calorie restriction) to initiate and maintain weight loss and improve liver health. Module 6 covers understanding food labels and portion sizes as well as practical strategies to overcome barriers to maintaining dietary behavior change. Module 7 focuses on the role of physical activity and exercise in the context of NAFLD, weight loss, and weight loss maintenance, with an emphasis on the type and amount of physical activity or exercise required to derive benefits on liver health. Finally, module 8 presents a suite of tools to facilitate self-regulation (eg, behavioral goal setting, feedback and self-monitoring tools for diet, physical activity, and weight). As a whole, the content of VITALISE incorporates recommendations made in national and international clinical guidelines for the management of patients with NAFLD [7,8,26].

Once enrolled in the study, patients received a welcome email with a unique link to access VITALISE using their personal email address. The information provided in each module is presented in a range of formats (eg, written text articles, interactive tools, and animations) to engage users and support encoding of information. Upon logging in, patients are asked to enter their height, weight, and weight loss goal. This information populates the self-regulation tools to facilitate tracking of progress and to generate automated feedback to support self-efficacy. Patients can upload details of their food choices (using photographs or a written description), daily step-count, and weight to enable self-monitoring.

Once participants had completed all 8 digital modules in a sequential order, they were granted access to a lifestyle coach. No specific recommendations on the frequency of contact with the coach were specified, although it was restricted to an initial 20 min appointment, followed by weekly 10 min appointments over the 12-week intervention period. The coaching sessions focused on details of participants' dietary choices, daily step counts or time spent active, and weight status to facilitate personalized goal setting and problem solving, with subsequent sessions focusing on review of behavioral goals and feedback on behavior.

Feedback on the prototype intervention from HCPs and patients in separate workshops provided evidence that the interactive components were useful and provided acceptable and accessible representations of what otherwise would have been complex information to communicate (eg, NAFLD progression). Feedback from both groups facilitated minor modifications to the written content of the intervention, including a clearer description of the noninvasive measures used to help identify disease severity in NAFLD, and the addition of links to healthy recipe ideas to support dietary behavior change.

Table 3. Overview of module content.

Modules and subcomponents	Description of information content (BCT[a] code numbers)
Module 1: Introduction	
Welcome to the program	Message from a specialist in lifestyle management of NAFLD[b] providing an overview of the aims and objectives of the program (9.1)
Module 2: Understanding NAFLD	
What is NAFLD?	Overview of the liver, what is NAFLD and how it progresses, and the different stages of the disease (4.2, 5.1)
How is NAFLD identified and diagnosed?	Overview of the tests and investigations that may be used to diagnose NAFLD (5.1, 5.2)
Understanding test results	What the different tests and investigations measure and what the results mean (5.1)
NAFLD myths	Common myths about NAFLD and why it should not be left unmanaged (4.2)
Managing your NAFLD	Animations demonstrating how lifestyle change can improve liver health and reduce cardiovascular disease risk and improve metabolic control (5.1, 5.2)
Patient success stories	Real-life examples of how people with NAFLD or at risk of NAFLD have improved their liver health by making lifestyle changes (6.2)
Long-term complications of NAFLD	Potential consequences of not managing NAFLD (5.1, 9.2)
Module 3: Getting started with coaching	How to access lifestyle coaches and what to expect from coaching sessions (4.1)
Module 4: Food and NAFLD	
Energy balance	How to balance energy requirements in the context of weight loss (2.7, 4.1, 8.2)
Nutrition and NAFLD	The relationship between diet quality, eating patterns, and NAFLD (4.2, 5.1, 5.2)
Understanding carbohydrates	What are carbohydrates and the role they play in diet (5.1)
Understanding fats	What are fats and the role they play in our diet (5.1)
Understanding alcohol	Recommended alcohol intake, calorie content in alcoholic drinks, and provision of a web-based alcohol calculator (2.3, 5.1)
Module 5: How do I make changes to my diet?	
Finding the right dietary approach	Animation explaining the evidence supporting 2 dietary approaches for managing NAFLD (5.2, 9.1)
Calorie restriction	Calorie restriction in more detail, provision of web-based weight loss calculators and calorie checkers, and link to the NHS[c] weight loss plan (2.7, 4.1, 5.1)
Mediterranean diet	Mediterranean diet in more detail (5.1)
Alternative dietary approaches	Information and external links to alternative dietary approaches (9.1)
Module 6: Practical tips	
Supermarket tour	Animation virtually walking the user around a supermarket providing practical tips on what to look out for when shopping and resisting temptation (1.2, 4.1)
Food labels	How to facilitate weight loss by understanding food labels (4.1)
Portion sizes	Recommended portion sizes of different food groups to promote weight loss and maintenance (4.1)
Healthy eating	Provides external links to healthy eating recipes to promote weight loss and maintenance (5.1, 9.1)
Module 7: Physical activity, exercise, and NAFLD	
How lifestyle can contribute to weight gain and NAFLD	Animation communicating how reduced physical activity levels are liked to weight gain and NAFLD (5.2)

Modules and subcomponents	Description of information content (BCT[a] code numbers)
Combining physical activity and diet to promote weight loss	The benefits of combining physical activity and dietary change to promote weight loss (5.1)
Physical activity, exercise, and NAFLD	The difference between physical activity and exercise and how physical activity and exercise can benefit the liver and lead to other health benefits (5.1)
How much physical activity do I need to do?	The required amount, intensity, and type of physical activity and exercise required to manage NAFLD and ideas for increasing physical activity and exercise levels (5.1)
Module 8: Steps to success	
5 essential steps to goal setting	Tips for setting SMART[d] goals (1.1)
Behavioral goal setting	Tool to facilitate personalized behavioral goal setting for diet and physical activity (1.1)
Dietary goal setting	Tool to facilitate monitoring of diet against dietary goals (2.3)
Step tracker	Tool to facilitate monitoring of daily step counts and track progress over time (2.3)
Weight tracking	Tool to facilitate weight monitoring and track progress (2.4)
Maximizing chances of success	Animation prompting use of SMART goals and self-monitoring and support from family and friends to maximize the chances of success (1.1, 1.4, 2.3, 2.4, 3.1, 3.2)
Lifestyle coaching support	Review behavioral goals (1.5), provide feedback on behavior (2.2), and provide social support (unspecified; 3.1) and social support (practical; 3.2). Social support is directed specifically at the target behaviors

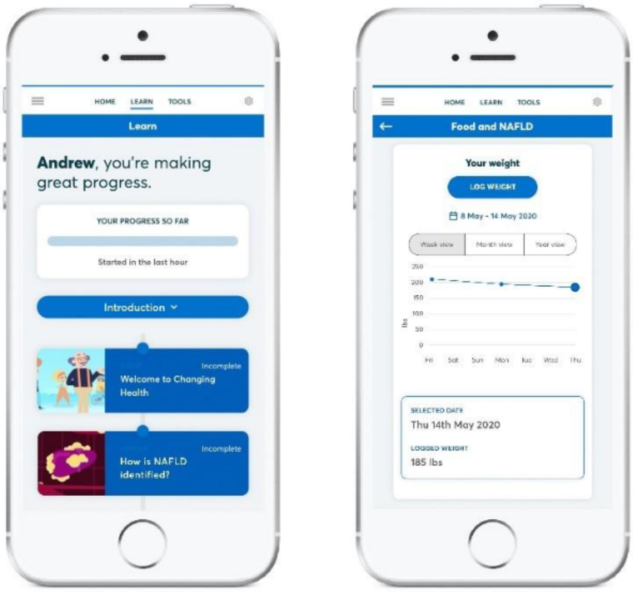
[a] BCT: behavior change technique.

[b] NAFLD: nonalcoholic fatty liver disease.

[c] NHS: National Health Service.

[d] SMART: Specific, Measurable, Achievable, Realistic, Timely.

Figure 2. Images of the digital interVention to promote lIfesTyle change in nonAlcoholic fatty LIver diseaSE (VITALISE).



Andrew, you're making great progress.

Started in the last hour

SELECTED DATE
Thu 14th May 2020

LOGGED WEIGHT
185 lbs

Intervention Mapping Step 5: Developing an Implementation Plan to Deliver VITALISE

Step 5 of the Intervention Mapping process involved working in collaboration with Changing Health Limited to develop an implementation plan. The plan consisted of training for lifestyle coaches, a standardized protocol for on-boarding (registration of patients) and making initial contact with patients, and allocation of coaching support.

A total of 3 lifestyle coaches employed by Changing Health Limited received training for the NAFLD prototype intervention. Training was delivered face-to-face in a single session by a senior physiotherapist with expertise in NAFLD and a health psychologist with expertise in lifestyle behavior change, and it lasted for 2 hours.

All 3 coaches were experienced in the delivery of lifestyle interventions in the context of long-term health conditions and providing remote support delivered by telephone. Therefore, the aims of the training session were to increase knowledge about NAFLD, including the management of NAFLD via lifestyle modification, and to further develop skills in delivering BCTs, specifically those incorporated within VITALISE (ie, strategies identified in step 3 to target motivation and enactment of behavioral change via engagement with the prototype digital intervention to promote weight loss and improve liver health).

The head of patient experience at Changing Health Limited agreed to make an initial contact with patients electing to take part in the program via email, providing them with a link to the digital intervention. If patients did not access the intervention within 7 days, a lifestyle coach assigned to VITALISE attempted to contact the patient via email and subsequently by telephone to prompt engagement. Changing Health agreed to provide patients with weekly lifestyle coaching sessions for 12 weeks during pilot testing of the intervention to allow them sufficient time and exposure to the intervention to provide feedback (step 6).

Intervention Mapping Step 6: Pilot Testing the Intervention

With reference to the implementation plan referred to in step 5, 16 patients with NAFLD who agreed to participate in the pilot were given access to VITALISE for 12 consecutive weeks. It is important to record the way in which participants accessed the intervention as well as the feedback they provided to inform future development and evaluation. A total of 11 of the 16 participants successfully logged on to the digital intervention and entered details of their height, weight, and weight loss goal. Out of the 16 participants, 5 experienced problems when attempting to log in, and despite receiving a reset email and prompting from a lifestyle coach to access the intervention, they disengaged from it.

Table 4 presents a summary of the findings for the 11 participants who continued to engage with VITALISE beyond their first log-in. Each participant accessed the intervention an average of 4 times over the piloting period. They accessed the module content an average of 7 times per log-in. Goal setting and self-monitoring tools were accessed on 22 occasions, and the 11 participants accessed these tools 4 times each on average. The modules accessed most frequently (between 7 and 10 unique accesses per participant) were *welcome to the program* (module 1), *understanding NAFLD* (module 2), and *food and NAFLD* (module 4). These modules were principally designed to target motivation for making lifestyle changes (behavioral intentions). The least frequently accessed modules covered topics to promote enactment of lifestyle behavior change, including goal setting, planning, problem solving, and monitoring. These findings supported the outcomes of the needs assessment conducted in step 1, where information about NAFLD was reported as most salient from the perspective of patients. Only 3 out of 11 participants requested access to a lifestyle coach for support throughout their use of the prototype intervention, and it was those participants who accessed the self-regulation tools (ie, goal setting and self-monitoring).

At the end of the 12-week piloting process, participants were emailed a brief qualitative questionnaire that asked them to provide written feedback on their experience of using the digital intervention. All 16 participants responded and provided feedback on the intervention, despite 5 being unable to access the intervention. When asked about the information content of the intervention, opinions varied between participants. For example, some reported the amount of information as too much, and at times difficult to understand, whereas others felt that the information provided was informative and pitched at the right level. The latter was particularly the case for participants who chose to access a lifestyle coach. A total of 2 participants indicated that they would usually prefer face-to-face support; however, they reported that the coaching support offered was innovative and an acceptable alternative. Those accessing a coach reported the support as beneficial for tailoring the information provided to their own individual circumstances. Participants reported that reviewing the information content motivated them to want to make lifestyle changes, but they did not provide any detail about why, for example, whether their primary motivator was liver health.

Access to VITALISE was reported to be the most salient barrier to engagement, and navigation for some was problematic, highlighting the need for further development, usability testing, and optimization. The participants who disengaged from the intervention did so due to several failed log-in attempts. As such, they reported losing confidence in the intervention despite being initially motivated to access it. It is important to note that the digital intervention was provided as a web link and not a mobile phone app. This was reported to affect initial motivation and engagement for those who expected to be able to access an app from the home screen of their mobile phone. This was reported to have reduced the number of times participants wanted to log-in (ie, ease of access would have promoted a higher level of use and engagement during the pilot). Those who disengaged and those who accessed the intervention reported this as having impacted their experience.

Access to the intervention was provided for 12 weeks; however, the majority of log-ins occurred within the first 4 weeks, where the information modules were accessed most frequently (ie, only 2 participants accessed the information module content on one occasion each after the initial 4 weeks, which corresponded to accessing a coach).

Table 4. Engagement with Intervention to Promote Lifestyle Change in Nonalcoholic Fatty Liver Disease by patients over a 12-week period.

Log-ins	Participants[a]
Distinct log-ins	
Total log-ins, n	57
Log-ins per person	
Mean (SD)	3.6 (3.5)
Range	0-14
Median (IQR)	4 (5)
Access to module content across 33 subsections	
Total views across all modules and subsections, n	208
Views per section	
Mean (SD)	6.3 (0.7)
Range	4-10
Median (IQR)	7 (4)
Access to goal setting and monitoring tools (5 individual tools available)	
Total log-ins, n	22
Log-ins across all 5 tools	
Mean (SD)	4.4 (0.7)
Range	3-6
Median (IQR)	4 (2)

[a] This refers to 11 out of 16 (69%) participants originally recruited who logged in to the intervention.

Discussion

Principal Findings

The first-line treatment for NAFLD is lifestyle behavior change to promote weight loss and liver health [4,5]. Nevertheless, behavior change strategies and structured education for NAFLD are not routinely used in clinical care [9]. This paper describes the systematic development of a co-designed, theory- and evidence-informed digital intervention—VITALISE—designed to target dietary and physical activity behaviors of adults with NAFLD, to initiate and maintain weight loss to improve liver health. VITALISE addresses the pressing need for a structured lifestyle program for people with NAFLD with evidence-based information and behavior change strategies with support from a lifestyle coach to maximize engagement.

Summary of Findings

Our qualitative study highlighted the need for information for patients about NAFLD, how it can progress, and how it can be managed. A minority of patients highlighted the need for support to make and sustain behavioral changes. Primary HCPs reported the need for information about NAFLD to raise their awareness and training to effectively communicate with patients about NAFLD and to provide support when required. Primary and secondary HCPs highlighted the importance of a referral pathway for patients to receive support to lose weight. They acknowledged a lack of skills in this area and a lack of time during consultations to effectively support patients to make behavioral changes [9]. Interview data were further analyzed using the TDF to identify targets for intervention. A total of 9 and 4 theoretical domains were identified in the context of HCP and patient interviews, respectively. These domains were used to select BCTs using a valid and reliable behavior change technique taxonomy [12].

Relation to Other Literature in the Field

Intervention mapping enabled the development of an intervention that addressed the specific needs and preferences of patients and HCPs identified, with content informed by theory and practical theory-linked behavioral strategies to engage patients in a change process at their own pace and stage of readiness. This is consistent with previous research evidence on the capability of digital interventions to provide an effective medium for targeting social cognitive determinants of behavior [19], including diet and physical activity [20].

The opportunity for patients to engage with lifestyle coaches by telephone has been found to increase engagement with digital interventions [21,22] and to overcome barriers to goal attainment for diet and physical activity behavior change [22]. The inclusion of specifically trained coaches in the use of lifestyle interventions to manage NAFLD was viewed as an essential, but underutilized, component of VITALISE. A total of 3 out of 11 patients accessed coaching support during the piloting of the intervention and were very positive about the experience. This indicates that some patients may prefer to access this additional support immediately as they are ready to make changes. Our experience with development of other digital interventions has indicated that only around 40% to 50% of patients want to access a coach at the commencement or early stages of an intervention

[22]. In the context of this study, it is possible that individuals accessed activities or support elsewhere or that they were not ready to make behavioral changes during the pilot phase. This needs to be explored further.

Modules 1, 2, and 4 were most often accessed, which could be explained by tunneling (delivery of the content in a prespecified order) when navigating the modules, although evidence has shown that the use of a free-roam (navigation in any order) did not improve engagement with or effectiveness of a digital brief alcohol intervention [27]. Therefore, it is possible that the tunneling of content and/or the amount of information provided by the digital intervention could have been overwhelming. Consequently, patients may not have persevered with the intervention long enough to gain access to a lifestyle coach. However, it is also possible that participants did not want to access a lifestyle coach and that information or education was sufficient for their needs. This assumption is congruent with the findings of our initial qualitative work [9], which was reported as part of step 1 of the development process.

Limitations

Although our findings are encouraging, they also suggest that further development and optimization is required to maximize engagement with VITALISE. The need to summarize some of the information provided in VITALISE, and identifying what is critical to access before gaining access to a lifestyle coach is warranted in future work. Specifically, further exploration is required to determine why most participants did not access coaching support. The intervention was made available to participants for 12 consecutive weeks, and this included access to coaching support (ie, 1 session per week for 12 weeks), and we acknowledge that this is likely to be insufficient to promote long-term behavior change. However, the aim of the pilot study was to obtain data on access and use of the intervention to inform optimization and not to assess changes in behavior.

Future Directions

The findings from this study have identified the need to further explore ways to maximize engagement with the intervention and utilization of coaching support. There is also a need to assess the feasibility of using the intervention within the clinical care pathway before assessing efficacy in a randomized controlled trial. A referral from an HCP may increase long-term engagement with the intervention when paired with positive reinforcement during annual review consultations [28], and this will be assessed. Future work will also involve the development and evaluation of a theory- and evidence-based training intervention for HCPs, which would accompany the patient intervention to support implementation at scale. On the basis of the findings from our qualitative work with HCPs, this is likely to include training for HCPs on how to provide information to patients on the importance of initiating lifestyle changes at the point of NAFLD diagnosis (to increase uptake of VITALISE), and personalized feedback for patients on the relationship between lifestyle behavior changes and clinical markers (eg, how changes in diet, physical activity, and weight loss impact blood test results [liver enzymes or glucose control] or specific liver investigations).

Although VITALISE was not piloted as part of clinical care delivery, it is easily embedded within the existing pathway of care. It provides HCPs in primary, secondary, and tertiary care centers with a mechanism to refer NAFLD patients for specialist behavior change support. This promotes more efficient use of clinicians' time, enabling them to focus on their core clinical roles. However, patients considered it important for HCPs to acknowledge any progress they had made using the digital program during review appointments, specifically about the impact of behavioral changes on clinical outcomes. This suggests a need to include a summary report for patients to take along to clinical appointments with their HCPs, or with permission from patients, HCPs are given access to an electronic summary sent to patient medical records. Additional benefits of the digital intervention are that it can be tailored to individual patient needs and is capable of being delivered at scale. Furthermore, it incorporates NAFLD clinical management guidelines and has the potential to be implemented for use as part of the NAFLD care pathway in both primary and secondary care.

Conclusions

In conclusion, intervention mapping provided a systematic user-centered method for developing a digital intervention with lifestyle coach support—VITALISE—for patients with NAFLD. The digital intervention was acceptable to patients and feasible to deliver. However, issues with initial access, optimization of information content, and promoting the value of remote support require further development ahead of future research to formally assess the acceptability and feasibility of the intervention in clinical settings and to establish the effectiveness of the intervention when delivered as part of existing care pathways.

Acknowledgments

KH is funded by a Clinical Lectureship (grant number CAT CL-2013-04-010) supported by the National Institute for Health Research and Health Education England. The views expressed in this manuscript are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health. The authors would like to acknowledge the design team at Changing Health Limited for providing their expertise to develop the digital intervention. The authors would also like to acknowledge the patients and HCPs who provided invaluable input and insights throughout the developmental process.

Authors' Contributions

KH, LA, SM, and QA conceived the study idea; LA designed the study; and KH and LA conducted the study and undertook data analysis. KH, LA, SM, DF, QA, and LH interpreted the findings to inform the developmental process. LA and KH drafted the initial manuscript. All study authors contributed to the writing and critical revision of the final manuscript for important intellectual content.

Conflicts of Interest

DF was previously employed by Changing Health Limited. LA has previously developed lifestyle behavior change interventions provided by Changing Health Limited.

References

1. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology 2018 Jan;67(1):123-133 [FREE Full text] [doi: 10.1002/hep.29466] [Medline: 28802062]
2. McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. J Hepatol 2015 May;62(5):1148-1155 [FREE Full text] [doi: 10.1016/j.jhep.2014.11.034] [Medline: 25477264]
3. Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat Rev Gastroenterol Hepatol 2013 Jun;10(6):330-344. [doi: 10.1038/nrgastro.2013.41] [Medline: 23507799]
4. Hallsworth K, Adams LA. Lifestyle modification in NAFLD/NASH: Facts and figures. JHEP Rep 2019 Dec;1(6):468-479 [FREE Full text] [doi: 10.1016/j.jhepr.2019.10.008] [Medline: 32039399]
5. Romero-Gómez M, Zelber-Sagi S, Trenell M. Treatment of NAFLD with diet, physical activity and exercise. J Hepatol 2017 Oct;67(4):829-846. [doi: 10.1016/j.jhep.2017.05.016] [Medline: 28545937]
6. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018 Jan;67(1):328-357. [doi: 10.1002/hep.29367] [Medline: 28714183]
7. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol 2016 Jun;64(6):1388-1402. [doi: 10.1016/j.jhep.2015.11.004] [Medline: 27062661]
8. Non-Alcoholic Fatty Liver Disease (NAFLD): Assessment and Management (NG49). National Institute for Clinical Excellence (NICE). 2016. URL: https://www.nice.org.uk/guidance/ng49 [accessed 2020-12-31]
9. Avery L, Exley C, McPherson S, Trenell MI, Anstee QM, Hallsworth K. Lifestyle behavior change in patients with nonalcoholic fatty liver disease: a qualitative study of clinical practice. Clin Gastroenterol Hepatol 2017 Dec;15(12):1968-1971. [doi: 10.1016/j.cgh.2017.06.011] [Medline: 28624648]
10. Bartholomew L, Parcel G, Kok G, Gottlieb N, Fernández M. Planning health promotion programs: An Intervention Mapping approachan Francisco, CA: Jossey-Bass. San Francisco,CA: Jossey-Bass; 2011.
11. Hallsworth K, Dombrowski SU, McPherson S, Anstee QM, Avery L. Using the theoretical domains framework to identify barriers and enabling factors to implementation of guidance for the diagnosis and management of nonalcoholic fatty liver disease: a qualitative study. Transl Behav Med 2020 Oct 8;10(4):1016-1030. [doi: 10.1093/tbm/ibz080] [Medline: 31120519]
12. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
13. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012;24(7):986.
14. Changing Health Ltd. URL: https://www.changinghealth.com/ [accessed 2020-11-09] [WebCite Cache ID https://www.changinghealth.com/]
15. Vilar-Gomez E, Martinez-Perez Y, Calzadilla-Bertot L, Torres-Gonzalez A, Gra-Oramas B, Gonzalez-Fabian L, et al. Weight loss through lifestyle modification significantly reduces features of nonalcoholic steatohepatitis. Gastroenterology 2015 Aug;149(2):367-78.e5; quiz e14. [doi: 10.1053/j.gastro.2015.04.005] [Medline: 25865049]
16. Promrat K, Kleiner DE, Niemeier HM, Jackvony E, Kearns M, Wands JR, et al. Randomized controlled trial testing the effects of weight loss on nonalcoholic steatohepatitis. Hepatology 2010 Jan;51(1):121-129 [FREE Full text] [doi: 10.1002/hep.23276] [Medline: 19827166]
17. Schwarzer R. Health action process approach (HAPA) as a theoretical framework to understand behavior change. AP 2016 Dec 5;30(121):119. [doi: 10.15517/ap.v30i121.23458]
18. van Achterberg T, Huisman-de Waal GG, Ketelaar NA, Oostendorp RA, Jacobs JE, Wollersheim HC. How to promote healthy behaviours in patients? An overview of evidence for behaviour change techniques. Health Promot Int 2011 Jun;26(2):148-162. [doi: 10.1093/heapro/daq050] [Medline: 20739325]
19. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]

20. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [doi: 10.1207/s15324796abm3103_2] [Medline: 16700634]
21. Little P, Stuart B, Hobbs FR, Kelly J, Smith ER, Bradbury KJ, et al. Randomised controlled trial and economic analysis of an internet-based weight management programme: POWeR+ (Positive Online Weight Reduction). Health Technol Assess 2017 Jan;21(4):1-62 [FREE Full text] [doi: 10.3310/hta21040] [Medline: 28122658]
22. Avery L, Flynn D, Ashley K, Trenell M. Using lifestyle coaches to promote use of self-regulation techniques within the mHealth intervention Changing Health. Eur Health Psychol 2018:-.
23. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
24. van DB, Schoones JW, Vliet VT. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]
25. Haigh L, Bremner S, Houghton D, Henderson E, Avery L, Hardy T, et al. Barriers and facilitators to mediterranean diet adoption by patients with nonalcoholic fatty liver disease in northern Europe. Clin Gastroenterol Hepatol 2019;17(7):1364-1371. [Medline: 30391437]
26. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology 2012 Jun;55(6):2005-2023. [doi: 10.1002/hep.25762] [Medline: 22488764]
27. Norman P, Webb TL, Millings A, Pechey L. Does the structure (tunneled vs free-roam) and content (if-then plans vs choosing strategies) of a brief online alcohol intervention effect engagement and effectiveness? A randomized controlled trial. Transl Behav Med 2019 Nov 25;9(6):1122-1130. [doi: 10.1093/tbm/ibz110] [Medline: 31287897]
28. Okwose NC, O'Brien N, Charman S, Cassidy S, Brodie D, Bailey K, et al. Overcoming barriers to engagement and adherence to a home-based physical activity intervention for patients with heart failure: a qualitative focus group study. BMJ Open 2020 Sep 21;10(9):e036382 [FREE Full text] [doi: 10.1136/bmjopen-2019-036382] [Medline: 32958484]

Abbreviations

BCT: behavior change technique
HAPA: health action process approach
HCP: health care professional
NAFLD: nonalcoholic fatty liver disease
NHS: National Health Service
TDF: Theoretical Domains Framework
VITALISE: Intervention to Promote Lifestyle Change in Nonalcoholic Fatty Liver Disease

Edited by G Eysenbach; submitted 18.06.20; peer-reviewed by C Greaves, K Knittle; comments to author 03.08.20; revised version received 05.10.20; accepted 28.10.20; published 15.01.21.

Please cite as:
Hallsworth K, McPherson S, Anstee QM, Flynn D, Haigh L, Avery L
Digital Intervention With Lifestyle Coach Support to Target Dietary and Physical Activity Behaviors of Adults With Nonalcoholic Fatty Liver Disease: Systematic Development Process of VITALISE Using Intervention Mapping
J Med Internet Res 2021;23(1):e20491
URL: http://www.jmir.org/2021/1/e20491/
doi:10.2196/20491
PMID:33448929

Original Paper

Efficacy of a Transdiagnostic Self-Help Internet Intervention for Reducing Depression, Anxiety, and Suicidal Ideation in Adults: Randomized Controlled Trial

Philip J Batterham[1], PhD; Alison L Calear[1], PhD; Louise Farrer[1], PhD; Amelia Gulliver[1], PhD; Ella Kurz[1], BMid

[1]Centre for Mental Health Research, Research School of Population Health, Australian National University, Canberra ACT, Australia

Corresponding Author:
Philip J Batterham, PhD
Centre for Mental Health Research
Research School of Population Health
Australian National University
63 Eggleston Road
Canberra ACT, 2601
Australia
Phone: 61 261251031
Email: philip.batterham@anu.edu.au

Abstract

Background: Low-intensity self-guided mental health interventions that are delivered on the web may meet the needs and preferences of adults with mild to moderate symptoms. However, few clinical trials have examined the effectiveness of self-guided transdiagnostic interventions within a naturalistic setting.

Objective: This randomized controlled trial (RCT) tests the effectiveness of the video-based transdiagnostic intervention *FitMindKit* in reducing depression symptoms (primary outcome), anxiety symptoms, disability, and suicidal ideation, relative to an attention-matched control condition called *HealthWatch*.

Methods: The RCT was conducted with adults living in the Australian Capital Territory, Australia. Participants (n=1986) were recruited through the web using social media advertisements, screened for psychological distress, and then randomized to receive one of two 4-week programs: *FitMindKit* (12-module psychotherapy intervention) or *HealthWatch* (12-module program providing general health information). Participants were assessed at baseline and at 4 weeks postbaseline. To maintain the ecological validity of the trial, participants completed brief assessments and interventions without direct researcher contact or incentives.

Results: Mixed model repeated-measures analyses of variance demonstrated that *FitMindKit* significantly improved depression symptoms ($F_{1,701.7}$=3.97; P=.047), along with panic symptoms ($F_{1,706.5}$=5.59; P=.02) and social anxiety symptoms ($F_{1,680.0}$=12.37; P<.001), relative to the attention control condition. There were no significant effects on other outcomes.

Conclusions: Self-guided transdiagnostic interventions can be beneficial when delivered directly to end users through the internet. Despite low adherence and small effect sizes, the availability of such interventions is likely to fill a critical gap in the accessibility of mental health services for the community.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618001688279; http://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=376113.

International Registered Report Identifier (IRRID): RR2-10.1016/j.conctc.2019.100341

(J Med Internet Res 2021;23(1):e22698) doi:10.2196/22698

KEYWORDS

depression; anxiety; randomized controlled trial; internet; implementation science

Introduction

Depression and anxiety are common mental disorders that account for significant global disease and disability burden [1]. Despite significant personal, interpersonal, social, educational, and vocational impacts of these disorders, only one in 3 people with a mental health problem seeks help from a health professional [2-4]. Those who engage in professional treatment typically delay seeking help for years or decades [5]. Cost of treatment, stigma, lack of knowledge, and low perceived need for treatment are commonly cited as barriers to seeking mental health help [6-8]. Internet-based, self-guided mental health interventions are able to mitigate some of these barriers and have been shown to be effective, efficient, and cost-effective in the prevention and treatment of many common mental disorders [9-11]. The high rate of comorbidity between depression, anxiety, and other mental health problems represents another significant challenge for help seeking and treatment, as it is often related to an increase in the severity and chronicity of symptoms [12,13]. One solution to the challenge of treating comorbid mental health problems is the use of transdiagnostic treatment approaches. Transdiagnostic treatments are typically based on the principles of cognitive behavioral therapy (CBT), which is an efficacious treatment approach for multiple mood and anxiety disorders [14,15]. Transdiagnostic treatments provide therapeutic content capable of targeting the common cognitive and behavioral processes that underlie the development and maintenance of multiple mental health disorders (eg, negative thinking patterns, avoidance, hyperarousal) and address symptoms that are common to multiple disorders (eg, sleep disturbance) [16,17].

Internet-based transdiagnostic mental health interventions have the potential to overcome barriers to mental health help seeking and treatment, and there is extensive evidence for their effectiveness [17-19]. There is evidence that such interventions may be more effective and efficient than interventions focusing on a single disorder [15]. However, the potential of web-based interventions to address comorbidity has not been fully realized. Trials of transdiagnostic internet interventions have largely been delivered within clinical settings [17] and, with very few exceptions [20], have focused on clinician-guided interventions. The provision of clinical support in an intervention requires increased resources, thereby reducing the potential for scalability. Furthermore, delivering interventions within a clinical setting neglects the majority of the population who do not engage in clinical care for mental health problems [2]. There is meta-analytic evidence that individuals with mild to moderate symptoms of depression or anxiety may benefit from self-guided interventions [21,22], which can be delivered without the need to connect with a health professional. Furthermore, delivery of self-guided internet interventions through web-based marketing may become an important component for prevention and early intervention, as it bypasses traditional treatment settings. Distal delivery of interventions may also become more critical in the face of public health crises and other system shocks, such as pandemics [23], and natural disasters (eg, bushfires, floods), when the delivery of traditional health services is hampered because of access restrictions.

As the abovementioned studies highlight, there is a need for randomized controlled trials (RCTs) comparing web-based interventions that address comorbidity with a robust attention control condition. As outlined in the protocol paper for this RCT [24], our research group has developed the *FitMindKit* program, a self-guided transdiagnostic web-based program that delivers CBT and other techniques through a series of brief videos and self-directed exercises. The program focuses on reducing both depression and anxiety symptoms using CBT-based content such as cognitive restructuring, problem-solving skills, and relaxation strategies. *FitMindKit* has been tested previously in an RCT comparing tailored and static versions of the program [25]. In that underpowered trial, there were no significant differences in effectiveness or adherence for a tailored version of the program, compared with a static version of the program. Consequently, this RCT tested a simplified static version of the program against an attention-matched control condition. This study was nested within a broader implementation study examining the uptake and reach of *FitMindKit* delivered directly to users through web-based advertising, in comparison with 2 community settings: primary care practices and pharmacies. The web-based arm of the implementation study is the focus of this paper, which used an RCT to test the effectiveness of the intervention in reducing depression symptoms.

We used a naturalistic RCT design for this study, where participants had no contact with researchers or clinicians and completed the intervention and brief assessments with only automated support (reminder emails). Although such an approach may limit adherence and increase attrition, leading to more modest effects [25-27], the methodology best reflects the delivery of self-help programs in the community [28]. Thus, the objective of this trial was to test the effectiveness of *FitMindKit* in reducing symptoms of depression and anxiety, suicidal ideation, and disability, when delivered in a real-world setting.

Methods

Trial Design

A two-arm parallel RCT was conducted comparing an active intervention (*FitMindKit)* with an attention control condition (*HealthWatch*). The trial protocol is provided in Batterham et al [24].

Ethics Approval

This trial was approved by the Australian National University Human Research Ethics Committee (ANU HREC protocol number 2017/911).

Participants and Procedure

Participants were recruited through the web between October 2018 and August 2019 via paid social media advertisements on Facebook and Instagram. Advertisements targeted adults aged 18 years and older living in the Australian Capital Territory to match the delivery catchment for the overarching implementation trial. On clicking on the Facebook advertisement, participants were directed to a webpage containing information about the study and questions to obtain their consent to participate. At this point, participants were assessed for eligibility as described in the following section, *Eligibility Criteria*. Following eligibility screening, a baseline assessment was completed, followed by randomization and treatment allocation, and a postintervention assessment following the 4-week intervention period.

Eligibility Criteria

To be eligible for the trial, participants were required to be aged 18 years and older and had moderate psychological distress as measured by the Distress Questionnaire-5 (DQ5) [29]. The DQ5 is a measure of psychological distress and consists of 5 items that assess the symptoms of common mental disorders. Participants were asked to endorse the frequency of each item over the last 30 days on a 5-point scale, ranging from never to always, with scores ranging from 5 to 25. In this study, the categories of DQ5 scores were no or low psychological distress (score of 5-7), moderate psychological distress/risk of a mental disorder (score of 8-17), and high-risk/probable clinical symptoms of a mental disorder (score of 18-25). These cut points were selected based on percentiles from existing population-based data [29]. Participants aged 18 years and older and who scored in the moderate risk category (score of 8-17) on the DQ5 were eligible for the trial, as a web-based self-guided mental health program is likely to be the most suitable for this group. Those in the low-risk category were provided with feedback and resources to access if their symptoms changed. Those in the high-risk category were strongly encouraged to seek help from a health professional and provided contact details for face-to-face, telephone-based, and web-based mental health resources and services.

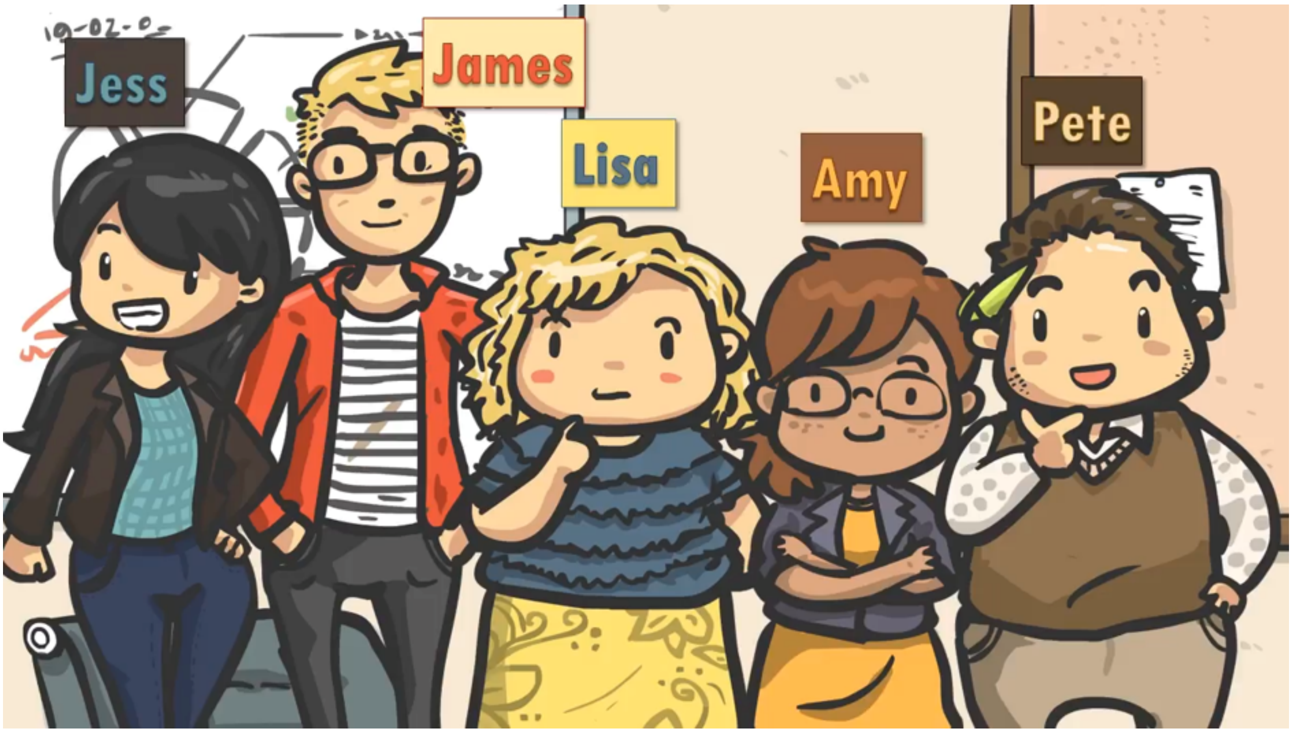
Eligible participants were invited to access the web-based trial portal containing trial assessments and intervention and control programs. Participants were required to create an account in the portal using an email address and password and then complete a brief baseline questionnaire. Following the completion of baseline measures, participants were automatically randomized, without involvement by trial staff, to receive the active intervention (*FitMindKit*) or the attention control program (*HealthWatch*). Randomization was in a 1:1 allocation ratio by a computer-generated sequence, with a block size of 6 stratified by age, sex, and DQ5 severity score, and the trial was double blinded. Following randomization, participants were provided with access to their assigned program and instructed to complete the program modules at their own pace over a 4-week period. They were also sent an email containing details of their allocated condition and a link to the portal. During the intervention period, participants received an automated weekly email reminding them to engage with their allocated program. Following the 4-week period, participants were sent an email inviting them to complete the postintervention survey. They received 2 reminder emails if they had not completed the survey after 1 and 2 weeks.

Interventions

Active Transdiagnostic Self-Guided Intervention: FitMindKit

Each *FitMindKit* module consists of a 2- to 6-minute video in which a series of fictional animated characters introduce concepts of and share their personal experience of one or more mental health problems. Figure 1 depicts *FitMindKit* characters. A relevant *homework* activity designed to facilitate practice of the therapeutic technique presented accompanies each video. Participants allocated to *FitMindKit* condition were able to access all 12 modules over the 4-week trial period. In total, 8 of the modules contained therapeutic techniques based on CBT principles (psychoeducation, getting help and support, cognitive reframing, problem solving, mindfulness, managing relationships, exercise and diet, and sleep hygiene), 2 modules targeted mood (behavioral activation and reducing rumination), 1 module targeted anxiety (exposure), and 1 module targeted suicidality (distress tolerance). Participants were directed to complete the module on psychoeducation first and were then free to choose when and in which order they completed the remaining modules over the 4-week period.

Figure 1. Screenshot of FitMindKit characters.



Attention Control Condition: HealthWatch

Participants in the attention control condition received a web-based health program called *HealthWatch*, a 12-module program pertaining to general health rather than mental health, which is not associated with therapeutic reductions in depression. Modules are text based and contain information on respiratory viruses, heart health, microbes, bone health, food hygiene, sun exposure, dietary supplements, kidney health, household burns, allergens, pancreas health, and posture. Participants randomized to the *HealthWatch* program were able to access the 12 modules over a 4-week period and complete them in any order.

Outcome Measures

The Patient Health Questionnaire-9 (PHQ-9) was the primary outcome measure, assessing the frequency of symptoms of major depression [30]. This scale consists of 9 items rated on a 4-point scale, ranging from *not at all* to *nearly every day*, and item scores are summed to produce an overall severity score ranging from 0 to 27, with higher scores indicating higher symptom severity. The PHQ-9 has sound sensitivity and specificity for detecting major depression in both clinical and general population samples and has been shown to detect change over time [31]. Internal consistency was adequate in the current sample at baseline (α=.78).

The Generalized Anxiety Disorder Scale-7 (GAD-7) was used to assess symptoms of generalized anxiety disorder (GAD) according to *Diagnostic and Statistical Manual of Mental Disorders* (DSM)-IV and DSM-5 diagnostic criteria [32]. The 7 items from the scale are rated on the same 4-point scale as the PHQ-9 and summed scores on the GAD-7 range from 0 to 21, with higher scores indicating greater symptom severity. Studies have demonstrated that the GAD-7 has good psychometric properties in general population samples [31,33]. Internal consistency was good in the current sample (α=.82).

Symptoms of panic and social anxiety were measured using the 4-item Panic Disorder Screener and the 4-item Social Anxiety Disorder Screener, respectively [34]. These scales have been developed and validated using Australian community–based samples, and they assess the frequency of panic symptoms (eg, “I had a sudden unexpected period of intense fear, anxiety or discomfort”) and social anxiety symptoms (eg, “I felt nervous during social situations”) in the past 30 days. Items are rated on a 5-point scale, ranging from *never* to *always*. Both screeners have demonstrated good convergent and divergent validity with diagnostic measures [34] and had good internal consistency in this sample (panic: α=.83; social anxiety: α=.91).

Suicidal ideation was measured using the Suicidal Ideation Attributes Scale (SIDAS) [35]. The SIDAS contains 5 items assessing the frequency of suicidal ideation, controllability of suicidal thoughts (reverse coded), closeness to suicide attempts, distress associated with suicidal thoughts, and functional impact of suicidal thoughts. Items are rated on a 10-point scale, and total scores range from 0 to 50, with scores greater than 21 indicating a high risk of suicidal behavior. The SIDAS has demonstrated high internal consistency and good convergent validity [35], with adequate internal consistency in this study sample at baseline (α=.76).

The extent of disability and disruption felt by participants because of mental health problems was measured with 2 items [36]: “How many days out of the past 30 were you totally unable to work or carry out your normal activities due to mental health problems?” (range 0-30) and “How many days out of the last 30 were you able to work or carry out your normal activities but had to cut back on what you did or did not get as much done as usual due to mental health problems?” (range 0-30). These items have been used previously in a large community-based epidemiological study of mental disorders in the United States [36].

Satisfaction was assessed using a 7-item scale assessing how much the participant (1) enjoyed the program, (2) found it helpful, (3) understood the content, (4) found it interesting, (5) would use it in the future, (6) would recommend it to others, and (7) “learnt” new skills from the program. Each item is rated on a 10-point scale, ranging from *completely disagree* to *completely agree*. The satisfaction scale had acceptable internal consistency in this study sample at baseline (α=.76).

Sample Size

The target sample size was 750, based on detecting a moderate effect size of Cohen *d* of 0.4 at posttest with 90% power, assuming 30% attrition from the trial. It was evident early in the trial that the rates of attrition were much higher than anticipated; however, demand for participation in the trial was also greater than anticipated. Consequently, we exceeded the original target to ensure that the final sample was greater than the 525 required at posttest, allowing us to detect more modest effects.

Statistical Methods

The effectiveness of the *FitMindKit* intervention was assessed using mixed model repeated-measures (MMRM) analyses of variance (ANOVA) [37] to compare mean scores on the outcome measures between the intervention and control groups. The critical test of effectiveness is the interaction between condition (intervention vs control) and time (posttest vs baseline). MMRM provides an intention-to-treat analysis, with unbiased estimates that account for all available data from participants who were enrolled in the trial [37]. An unstructured variance-covariance matrix was assumed, and df were estimated with Satterthwaite correction. The primary outcome was depression symptoms (PHQ-9) at the posttest end point. Alpha was set at $P<.05$ for MMRM analyses. Evidence for moderation of outcomes by gender, age group, educational attainment, and module completion was tested using MMRM analyses with three-way interactions of moderator×time×condition. Alpha was set at $P<.01$ for moderation analyses to account for multiple comparisons. Finally, satisfaction was compared across the 2 conditions on the basis of a two-tailed *t* test.

Results

A total of 1986 adults were recruited and randomized into the study between October 2018 and September 2019. Table 1 presents the characteristics of the sample by trial condition. The sample predominantly comprised females aged between 26 and 55 years, who were employed full-time, and from an English-speaking background. Figure 2 depicts the CONSORT (Consolidated Standards of Reporting Trials) flow diagram for the trial. Considerable attrition from the trial was observed, with only 34.19% (679/1986) of participants across both conditions completing the posttest assessment. Although older participants were more likely to complete the posttest (χ^2_5=32.4; P<.001), there were no differences in attrition by gender (χ^2_3=0.5; P=.93) or education (χ^2_9=12.9; P=.17). Similarly, there were no differences in attrition by any of the baseline symptom measures (P>.05), except for a difference in suicidal ideation severity (completers: mean 7.09, SD 9.04; dropouts: mean 6.12, SD 8.54; t_{1826}=2.24; P=.03). The proportion of participants who completed the posttest assessment was significantly higher in the control condition (391/995, 39.5%) than in the intervention condition (288/991, 28.9%; P<.001). Intervention adherence was also low—participants accessed a mean of 2.2 (SD 3.4) of the 12 *FitMindKit* modules, with 13.4% (133/991) completing 6 or more modules. In contrast, participants in the HealthWatch control condition completed a mean of 6.2 (SD 5.1) modules. There were no differences in demographic or outcome variables between conditions at baseline.

Table 1. Sample characteristics based on condition.

Characteristic	FitMindKit (intervention; n=991), n (%)	HealthWatch (attention control; n=995), n (%)	Chi-square (*df*)	*P* value
Gender			0.4 (3)	.95
Male	141 (14.2)	136 (13.7)		
Female	841 (84.9)	850 (85.4)		
Other	3 (0.3)	4 (0.4)		
Prefer not to answer	6 (0.6)	5 (0.5)		
Age group (years)			8.0 (5)	.16
18-25	134 (13.5)	134 (13.5)		
26-35	279 (28.2)	232 (23.3)		
36-45	214 (21.6)	219 (22.0)		
46-55	204 (20.6)	223 (22.4)		
56-65	121 (12.2)	134 (13.5)		
66+	39 (3.9)	53 (5.3)		
Education			1.0 (4)	.92
High school or less	191 (19.3)	184 (18.5)		
Certificate, diploma, or associate degree	242 (24.4)	258 (25.9)		
Bachelor’s degree	260 (26.2)	265 (26.6)		
Higher degree	294 (29.7)	285 (28.6)		
No answer	4 (0.4)	3 (0.3)		
Employment status			2.6 (4)	.63
Full-time	519 (52.4)	494 (49.6)		
Part-time/casual	255 (25.7)	257 (25.8)		
Unemployed	69 (7.0)	82 (8.2)		
Not working	133 (13.4)	143 (14.4)		
Prefer not to answer	15 (1.5)	19 (1.9)		
Language spoken at home			0.7 (1)	.42
English only	902 (91.0)	895 (89.9)		
Other	89 (9.0)	100 (10.1)		
Completed posttest			23.1 (1)	<.001
Yes	288 (28.9)	391 (39.5)		
No	703 (70.7)	604 (60.9)		
Modules accessed			342.0 (4)	<.001
0	418 (42.2)	181 (18.2)		
1-2	329 (33.2)	213 (21.4)		
3-6	133 (13.4)	140 (14.1)		
7-11	41 (4.1)	92 (9.2)		
12	70 (7.1)	369 (37.1)		

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. DQ5: Distress Questionnaire-5; iCBT: internet-based cognitive behavioral therapy.

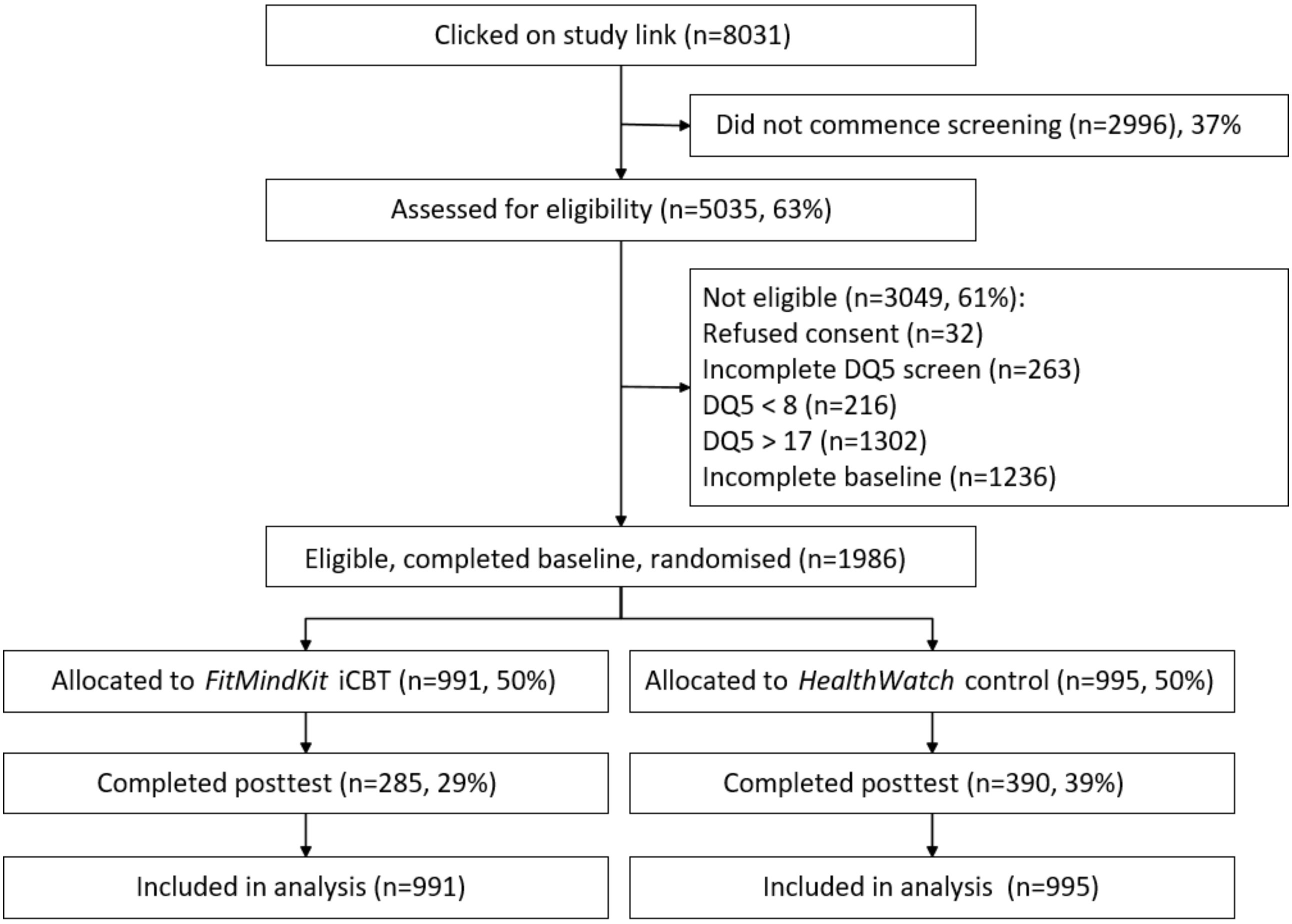


Table 2 provides the observed means and SDs for the outcome variables across the 2 conditions, and Table 3 presents the estimates of fixed effects from the MMRM ANOVA models. There was a significant time-by-condition interaction effect for the primary outcome of PHQ-9 depression scores, indicating that participants in the intervention condition had a significantly greater reduction in depression symptoms from baseline to posttest relative to the control condition ($F_{1,701.7}$=3.97; P=.047). For the secondary outcomes, there were also significant effects found for the intervention on symptoms of panic (Panic Disorder Screener $F_{1, 706.5}$=5.59; P=.02) and social anxiety (Social Anxiety Disorder Screener $F_{1,680.0}$=12.37; P<.001). Between-group effect sizes were very small for depression (d=0.06) and panic (d=0.08) symptoms and small for social anxiety symptoms (d=0.16). Although there were significant reductions over time for both conditions in symptoms of GAD and suicidal ideation, there were no significant differences between the 2 conditions from pretest to posttest. Likewise, the intervention did not have a significant effect on days out of role. The moderation analyses found no significant evidence that intervention effects were different on the basis of age, gender, educational attainment, or number of modules accessed, for any of the outcomes. Finally, overall satisfaction scores with the intervention at posttest were significantly higher in *FitMindKit* condition (mean 6.5, SD 2.0) than in the *HealthWatch* control condition (mean 5.3, SD 2.2; $t_{589.1}$=7.0; P<.001).

Table 2. Observed means (SD) for outcome variables at baseline and posttest.

Outcome	FitMindKit (intervention)				HealthWatch (attention control)			
	Baseline		Posttest		Baseline		Posttest	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
Patient Health Questionnaire-9: depression	989	10.60 (4.95)	285	8.82 (5.68)	988	10.38 (4.93)	390	8.92 (5.57)
Generalized Anxiety Disorder Scale-7: anxiety	989	8.16 (4.38)	280	6.71 (4.70)	993	8.08 (4.41)	381	6.90 (4.72)
Panic Disorder Screener: panic	988	3.20 (2.97)	280	2.62 (2.93)	993	3.10 (2.81)	380	2.74 (3.04)
Social Phobia Screener: social anxiety	991	6.17 (3.80)	280	5.23 (4.10)	994	6.10 (3.75)	380	5.75 (4.26)
Suicidal Ideation Attributes Scale: suicidal ideation	917	6.47 (8.69)	267	5.88 (8.36)	911	6.44 (8.77)	357	5.65 (8.30)
Days out of role	991	2.77 (5.47)	280	2.69 (5.70)	995	2.72 (5.38)	379	2.45 (5.68)

Table 3. Estimates of fixed effects from mixed model repeated-measures models.

Outcome and source	*F* test (*df*)	*P* value
Patient Health Questionnaire-9 : depression score		
Intercept	5768.89 (1,1424.1)	<.001
Condition	0.28 (1,1424.1)	.60
Time	78.23 (1,701.7)	<.001
Condition×time	3.97 (1,701.7)	.047
Generalized Anxiety Disorder Scale-7 : generalized anxiety score		
Intercept	4607.66 (1,1366.3)	<.001
Condition	0.51 (1,1366.3)	.48
Time	66.79 (1,715.4)	<.001
Condition×time	2.21 (1,715.4)	.14
Panic Disorder Screener : panic score		
Intercept	1755.96 (1,1556.0)	<.001
Condition	0.72 (1,1556.0)	.40
Time	26.87 (1,706.5)	<.001
Condition×time	5.59 (1,706.5)	.02
Social Phobia Screener : social anxiety score		
Intercept	3717.83 (1,1613.4)	<.001
Condition	3.14 (1,1613.4)	.08
Time	40.50 (1,680.0)	<.001
Condition×time	12.37 (1,680.0)	<.001
Suicidal Ideation Attributes Scale : suicidal ideation score		
Intercept	844.95 (1,1501.4)	<.001
Condition	0.02 (1,1501.4)	.88
Time	11.80 (1,687.3)	.001
Condition×time	0.06 (1,687.3)	.81
Days out of role		
Intercept	388.09 (1,1458.1)	<.001
Condition	0.06 (1,1458.1)	.80
Time	2.05 (1,718.7)	.15
Condition×time	0.42 (1,718.7)	.52

Discussion

Principal Findings

This study demonstrated that a self-guided transdiagnostic intervention was effective in reducing symptoms of depression, panic, and social anxiety when delivered directly to end users in a naturalistic setting. The effects of the intervention relative to the control condition were modest and did not extend to symptoms of generalized anxiety or suicidal ideation. This outcome is not surprising, as engagement with the trial and adherence to the intervention were both low, despite participants reporting greater satisfaction with the active intervention relative to the control. The absence of human support typically leads to lower levels of engagement in clinical trials [38-40], leading to attrition levels that are more commensurate with the naturalistic delivery of the intervention in the community than with an in-person clinical trial [28]. A previous review has also emphasized that attrition is typically higher for active interventions than control conditions because of factors such as higher expectations for change and greater cognitive demands [41].

Despite low adherence, this approach of delivering low-intensity self-guided interventions is highly scalable, as the resources required to deliver the intervention to the general community on the basis of demand are minimal. Assuming that 20% of adults experience elevated depression or anxiety symptoms [42], the trial reached more than 3% of the target population over 10 months, which would be challenging to achieve through other approaches such as recruitment through health services. The investment of time for users of the intervention was minimal, with participants completing a mean of 2.2 modules, which equates to approximately 6 to 12 minutes of engagement. Although few participants received a high dose of the intervention, it is clear that a proportion of users who received the intervention demonstrated significant benefits in terms of symptom reduction compared with those in the attention control condition. Although the differences between conditions on symptoms of mental health problems were modest at an individual level, when scaled to a population level, such modest effects are likely to have a significant impact. In addition, the high ecological validity of the trial suggests that the results are likely to be reflected outside of the research context, where programs are delivered in community settings, including directly to consumers through web-based marketing.

Implications

Given the demand for the program, the evidence of benefit, and the limited resources required to deliver self-guided transdiagnostic interventions such as *FitMindKit*, there appears to be a case for making such programs publicly available, with potential for impact at a population level when delivered at scale. In particular, individuals with mild to moderate symptoms who are reluctant to engage with mental health services but recognize a need for support may benefit from self-guided transdiagnostic interventions [40]. Similarly, individuals with limited mental health literacy or high levels of mental illness stigma may find that such interventions overcome these barriers [43]. Self-guided interventions may act as a gateway to more formal psychological services, providing users with exposure to the kinds of strategies they would likely encounter when engaging with a mental health professional [40,44]. This is particularly the case for accessible programs such as *FitMindKit* that use video and brief modules focused on the core elements of CBT delivered in a flexible way determined by the user, in contrast to programs that adapt the 50-minute psychological consultation model into a web-based setting. Transdiagnostic internet interventions may fill an important gap in the delivery of evidence-based services, particularly in regions that are rural or economically disadvantaged, where traditional services may be difficult to access [45].

The differential effects of the intervention for different types of mental health symptoms may reflect the features of the intervention or the types of people who engage in internet-based programs. The strongest effects were seen for social anxiety symptoms, which may reflect the considerable treatment delays seen among people with social anxiety disorder, which typically span 16 years [5]. That is, internet-based self-help interventions may be the only evidence-based support that many people with social anxiety are willing to access. The lack of effect for GAD symptoms may suggest that additional content focused on worry in addition to rumination may be beneficial. Suicidal ideation may be more difficult to address using a transdiagnostic intervention without a specific focus on suicide and distress reducing strategies and may require a higher level of support in the context of internet interventions [35,46]. Functional impairment was also not significantly decreased by the intervention, although it is possible that the time scale for reducing functional outcomes may be lengthy, requiring longer-term follow-up.

Strengths and Limitations

This study was one of the first to test the efficacy of a self-guided transdiagnostic web-based intervention. The strengths of the study include the large population-based sample, rigorous design, and high ecological validity of the methodology. However, the ecological validity was also a weakness because of the high rates of attrition and low adherence. Adherence was lower in the active condition than in the control condition, which may suggest that the control program may have been less cognitively demanding. Alternatively, the low adherence in the CBT intervention may reflect the challenges of engaging in video content. The analytical method used was robust to attrition, assuming data were missing at random (ie, missing on the basis of the observed characteristics of the sample). Previous research has indicated that differential attrition has little impact on the estimate of intervention effects [47], particularly when unbiased statistical methods are used that account for all available data [37]. Nevertheless, further evaluation of the intervention's efficacy and exploration of the factors that support its uptake may be warranted. The sample was not fully representative of the population of interest, with an overrepresentation of females in particular. Although females experience a higher prevalence of mental health conditions, the underrepresentation of males may reflect ongoing challenges in engaging males in therapy, particularly in digital interventions [48]. By necessity, outcomes were assessed using brief but accurate self-report measures, which may not reflect specific clinical outcomes. The naturalistic design of the study precluded medium- or long-term follow-up, so the duration of impact could not be determined. Finally, the effect sizes were very small, suggesting that the intervention may have had minimal clinical impact for many users. Future research may benefit from investigating subgroups of the population who are most suited to brief self-guided interventions.

Conclusions

A self-guided transdiagnostic intervention delivered through the internet without human contact was effective in reducing mental health symptoms. Although adherence was low and the effects of the intervention were modest, this study demonstrates the potential public health benefits of delivering low-intensity mental health programs at scale directly to consumers through the internet.

Acknowledgments

This research was funded by a grant from the National Health and Medical Research Council (NHMRC, Australia; 1142363). The study is related to larger program of research funded by European Union Horizons 2020 Project 733025, ImpleMentAll, although the study received no funding from the European Union. PB was supported by an NHMRC fellowship (1158707). AC was supported by NHMRC fellowships 1122544 and 1173146. LF was supported by an Australian Research Council Discovery Early Career Research Award (DE190101382).

Authors' Contributions

PB and AC designed the study and obtained the project funding. LF and EK implemented the study. PB analyzed the data. EK and PB drafted the manuscript. All authors provided critical revisions to the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1245 KB - jmir_v23i1e22698_app1.pdf]

References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. The Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/s2215-0366(15)00505-2]
2. Burgess PM, Pirkis JE, Slade TN, Johnston AK, Meadows GN, Gunn JM. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2009 Jul;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
3. Fried EI, Nesse RM. The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS One 2014 Feb 28;9(2):e90311 [FREE Full text] [doi: 10.1371/journal.pone.0090311] [Medline: 24587318]
4. Kessler RC, Heeringa S, Lakoma MD, Petukhova M, Rupp AE, Schoenbaum M, et al. Individual and societal effects of mental disorders on earnings in the United States: results from the national comorbidity survey replication. Am J Psychiatry 2008 Jun;165(6):703-711 [FREE Full text] [doi: 10.1176/appi.ajp.2008.08010126] [Medline: 18463104]
5. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun 01;62(6):603-613. [doi: 10.1001/archpsyc.62.6.603] [Medline: 15939838]
6. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol. Med 2013 Aug 09;44(6):1303-1307. [doi: 10.1017/s0033291713001943]
7. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10(1). [doi: 10.1186/1471-244x-10-113]
8. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol. Med 2010 Dec 07;41(8):1751-1761. [doi: 10.1017/s0033291710002291]
9. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016 Mar 28;12(1):157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
10. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 07;192(S11):S4-11. [doi: 10.5694/j.1326-5377.2010.tb03685.x] [Medline: 20528707]
11. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]

12. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun 01;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
13. Hirschfeld RMA. The comorbidity of major depression and anxiety disorders: recognition and management in primary care. Prim Care Companion J Clin Psychiatry 2001 Dec 01;3(6):244-254 [FREE Full text] [doi: 10.4088/pcc.v03n0609] [Medline: 15014592]
14. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb 04;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
15. Newby JM, Mewton L, Andrews G. Transdiagnostic versus disorder-specific internet-delivered cognitive behaviour therapy for anxiety and depression in primary care. J Anxiety Disord 2017 Mar;46:25-34. [doi: 10.1016/j.janxdis.2016.06.002] [Medline: 27396841]
16. Harvey A. Cognitive behavioural processes across psychological disorders : a transdiagnostic approach to research and treatment. In: Oxford Clinical Psychology. New York: Oxford University Press; 2004:a.
17. Newby JM, Twomey C, Yuan Li SS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
18. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
19. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan 07;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
20. Zwerenz R, Becker J, Johansson R, Frederick RJ, Andersson G, Beutel ME. Transdiagnostic, psychodynamic web-based self-help intervention following inpatient psychotherapy: results of a feasibility study and randomized controlled trial. JMIR Ment Health 2017 Oct 16;4(4):e41 [FREE Full text] [doi: 10.2196/mental.7889] [Medline: 29038094]
21. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
22. Mayo Wilson E, Montgomery P. Media‐delivered cognitive behavioural therapy and behavioural therapy (self‐help) for anxiety disorders in adults. Cochrane Database of Systematic Reviews 2013 Sep 09. [doi: 10.1002/14651858.cd005330.pub4]
23. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
24. Batterham PJ, Calear AL, Gulliver A, Farrer LM. Efficacy of a transdiagnostic, video-based online program for reducing depression, anxiety, and suicidal ideation in adults: Protocol for a randomised controlled trial. Contemp Clin Trials Commun 2019 Jun;14:100341 [FREE Full text] [doi: 10.1016/j.conctc.2019.100341] [Medline: 30911696]
25. Batterham PJ, Calear AL, Farrer L, McCallum SM, Cheng VWS. FitMindKit: randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interv 2018 Jun;12:91-99 [FREE Full text] [doi: 10.1016/j.invent.2017.08.002] [Medline: 30135773]
26. Batterham PJ, van Spijker BAJ, Mackinnon AJ, Calear AL, Wong Q, Christensen H. Consistency of trajectories of suicidal ideation and depression symptoms: evidence from a randomized controlled trial. Depress Anxiety 2019 Apr 22;36(4):321-329. [doi: 10.1002/da.22863] [Medline: 30466148]
27. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. The Lancet Psychiatry 2016 Apr;3(4):333-341. [doi: 10.1016/s2215-0366(15)00536-2]
28. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]
29. Batterham PJ, Sunderland M, Carragher N, Calear AL, Mackinnon AJ, Slade T. The Distress Questionnaire-5: population screener for psychological distress was more accurate than the K6/K10. J Clin Epidemiol 2016 Mar;71:35-42. [doi: 10.1016/j.jclinepi.2015.10.005] [Medline: 26464194]
30. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
31. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
32. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

33. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
34. Sunderland M, Batterham PJ, Calear AL, Carragher N. The development and validation of static and adaptive screeners to measure the severity of panic disorder, social anxiety disorder, and obsessive compulsive disorder. Int J Methods Psychiatr Res 2017 Dec;26(4):e1561 [FREE Full text] [doi: 10.1002/mpr.1561] [Medline: 28370638]
35. van Spijker BAJ, Batterham PJ, Calear AL, Farrer L, Christensen H, Reynolds J, et al. The suicidal ideation attributes scale (SIDAS): community-based validation study of a new scale for the measurement of suicidal ideation. Suicide Life Threat Behav 2014 Aug;44(4):408-419. [doi: 10.1111/sltb.12084] [Medline: 24612048]
36. Kessler RC, Frank RG. The impact of psychiatric disorders on work loss days. Psychol Med 1997 Jul;27(4):861-873. [doi: 10.1017/s0033291797004807] [Medline: 9234464]
37. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. In: Probability Theory and Stochastic Processes. New York: Springer-Verlag; 2000.
38. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
39. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
40. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014 Nov;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
41. Crutzen R, Viechtbauer W, Spigt M, Kotz D. Differential attrition in health behaviour change trials: a systematic review and meta-analysis. Psychol Health 2015 Jan;30(1):122-134. [doi: 10.1080/08870446.2014.953526] [Medline: 25109224]
42. Slade T, Johnston A, Oakley Browne MA, Andrews G, Whiteford H. 2007 National Survey of Mental Health and Wellbeing: methods and key findings. Aust N Z J Psychiatry 2009 Jul;43(7):594-605. [doi: 10.1080/00048670902970882] [Medline: 19530016]
43. Barak A, Grohol JM. Current and future trends in internet-supported mental health interventions. J of Tech in Human Ser 2011 Jul;29(3):155-196. [doi: 10.1080/15228835.2011.616939]
44. Titov N. Status of computerized cognitive behavioural therapy for adults. Aust N Z J Psychiatry 2007 Feb;41(2):95-114. [doi: 10.1080/00048670601109873] [Medline: 17464688]
45. Turpin M, Bartlett H, Kavanagh D, Gallois C. Mental health issues and resources in rural and regional communities: an exploration of perceptions of service providers. Aust J Rural Health 2007 Apr;15(2):131-136. [doi: 10.1111/j.1440-1584.2007.00870.x] [Medline: 17441823]
46. van Spijker BA, Werner-Seidler A, Batterham PJ, Mackinnon A, Calear AL, Gosling JA, et al. Effectiveness of a web-based self-help program for suicidal thinking in an australian community sample: randomized controlled trial. J Med Internet Res 2018 Feb 14;20(2):e15 [FREE Full text] [doi: 10.2196/jmir.8595] [Medline: 29444769]
47. Hewitt CE, Kumaravel B, Dumville JC, Torgerson DJ, Trial attrition study group. Assessing the impact of attrition in randomized controlled trials. J Clin Epidemiol 2010 Nov;63(11):1264-1270. [doi: 10.1016/j.jclinepi.2010.01.010] [Medline: 20573482]
48. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
DQ5: Distress Questionnaire-5
DSM: Diagnostic and Statistical Manual of Mental Disorders
GAD: generalized anxiety disorder
GAD-7: Generalized Anxiety Disorder Scale-7
MMRM: mixed model repeated measures
NHMRC: National Health and Medical Research Council
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
SIDAS: Suicidal Ideation Attributes Scale

Edited by G Eysenbach; submitted 21.07.20; peer-reviewed by D Wang, A AL-Asadi; comments to author 31.08.20; revised version received 21.09.20; accepted 25.10.20; published 22.01.21.

Please cite as:
Batterham PJ, Calear AL, Farrer L, Gulliver A, Kurz E
Efficacy of a Transdiagnostic Self-Help Internet Intervention for Reducing Depression, Anxiety, and Suicidal Ideation in Adults: Randomized Controlled Trial
J Med Internet Res 2021;23(1):e22698
URL: https://www.jmir.org/2021/1/e22698
doi:10.2196/22698
PMID:33480860

Original Paper

Effectiveness of Smartphone-Based Mindfulness Training on Maternal Perinatal Depression: Randomized Controlled Trial

Yaoyao Sun[1], BMed; Yanyan Li[1], BMed; Juan Wang[1], BMed; Qingyi Chen[1], BMed; Alessandra N Bazzano[2], MPH, PhD; Fenglin Cao[1], BMed, MM, PhD

[1]School of Nursing and Rehabilitation, Shandong University, Jinan, China

[2]Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, United States

Corresponding Author:
Fenglin Cao, BMed, MM, PhD
School of Nursing and Rehabilitation
Shandong University
44 Wenhua Xi Road
Jinan, 250012
China
Phone: 86 053188382291
Email: caofenglin2008@126.com

Abstract

Background: Despite potential for benefit, mindfulness remains an emergent area in perinatal mental health care, and evidence of smartphone-based mindfulness training for perinatal depression is especially limited.

Objective: The objective of this study was to evaluate the effectiveness of a smartphone-based mindfulness training intervention during pregnancy on perinatal depression and other mental health problems with a randomized controlled design.

Methods: Pregnant adult women who were potentially at risk of perinatal depression were recruited from an obstetrics clinic and randomized to a self-guided 8-week smartphone-based mindfulness training during pregnancy group or attention control group. Mental health indicators were surveyed over five time points through the postpartum period by online self-assessment. The assessor who collected the follow-up data was blind to the assignment. The primary outcome was depression as measured by symptoms, and secondary outcomes were anxiety, stress, affect, sleep, fatigue, memory, and fear.

Results: A total of 168 participants were randomly allocated to the mindfulness training (n=84) or attention control (n=84) group. The overall dropout rate was 34.5%, and 52.4% of the participants completed the intervention. Mindfulness training participants reported significant improvement of depression (group × time interaction χ^2_4=16.2, *P*=.003) and secondary outcomes (χ^2_4=13.1, *P*=.01 for anxiety; χ^2_4=8.4, *P*=.04 for positive affect) compared to attention control group participants. Medium between-group effect sizes were found on depression and positive affect at postintervention, and on anxiety in late pregnancy (Cohen d=0.47, –0.49, and 0.46, respectively). Mindfulness training participants reported a decreased risk of positive depressive symptom (Edinburgh Postnatal Depression Scale [EPDS] score>9) compared to attention control participants postintervention (odds ratio [OR] 0.391, 95% CI 0.164-0.930) and significantly higher depression symptom remission with different EPDS reduction scores from preintervention to postintervention (OR 3.471-27.986). Parity did not show a significant moderating effect; however, for nulliparous women, mindfulness training participants had significantly improved depression symptoms compared to nulliparous attention control group participants (group × time interaction χ^2_4=18.1, *P*=.001).

Conclusions: Smartphone-based mindfulness training is an effective intervention in improving maternal perinatal depression for those who are potentially at risk of perinatal depression in early pregnancy. Nulliparous women are a promising subgroup who may benefit more from mindfulness training.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900028521; http://www.chictr.org.cn/showproj.aspx?proj=33474

(J Med Internet Res 2021;23(1):e23410) doi:10.2196/23410

KEYWORDS

mindfulness; pregnancy; perinatal care; depression; mHealth; mobile health; psychosocial intervention

Introduction

The perinatal period is a major transitional stage in the life course, accompanied by social, emotional, physical, and hormonal changes, and is consequently an important period for mental health problems [1]. Symptoms of perinatal psychological distress are experienced by a large number of women with approximately 5%-30% affected by depressive symptoms [2]. Perinatal depression, a common nonpsychotic depressive disorder, occurs through pregnancy to the postpartum phase, and has serious consequences for maternal [3], infant, and childhood outcomes [4], as well as economic costs [5]. Low-and middle-income countries (LAMICs) reportedly have a higher prevalence of prenatal depression than high-income countries (19%-25% vs 7%-15%) [4]. However, the prevention and treatment of perinatal depression in LAMICs remains underrecognized, in part due to limited resources for mental health services, greater priority placed on preventing obstetric complications and fetal anomaly [6], and fear of stigmatization [7]. In this context, provision of low-cost and effective mental health services in LAMICs represents a potentially significant strategy to improve maternal and child health over the life course.

Difficulty in accessing health services has been cited as a reason for the low uptake of mental health interventions in previous studies [8,9]. The pervasive availability of smartphones provides an appropriate platform to address this problem and improve accessibility to mental health services [10]. Smartphone apps were reported to be helpful for treating depression in a recent systematic review and meta-analysis [11]. Another meta-analysis of randomized controlled trials (RCTs) including 22 apps also illustrated the promise of smartphones as a self-management tool for depression, with effect sizes ranging from 0.22 to 0.56 [12]. Smartphone-based interventions may be particularly useful and applicable for perinatal populations because recent studies have found that a large proportion of women seek pregnancy-related information through the internet [13,14].

Among the popular mental health smartphone apps, mindfulness was considered the most common evidence-based strategy [10] and one of the most frequently used modalities in apps for depression treatment [15]. Mindfulness is defined as “paying attention in a particular way: on purpose, in the present moment, and non-judgmentally” [16], and contains seven fundamental attitudinal factors: nonjudging, nonstriving, beginner’s mind, patience, trust, acceptance, and letting go [17]. Mindfulness-based interventions (MBIs) have now been widely used for reducing depressive symptoms, relieving psychological distress, and improving wellness within a broad range of populations [18]. For pregnant women, prenatal MBIs such as the Mindfulness-Based Childbirth and Parenting [19] and MindBabyBody [20] programs have been utilized to reduce maternal depression, anxiety, and negative affect [21]; enhance maternal nurturing behaviors; and improve childhood outcomes [22]. MBIs support individuals to alter intrinsic thought patterns, explore mind-body connections, and develop behavior modifications [23], which could be particularly appropriate for pregnant women facing physical changes and social role adjustment. Some reviews have reported the increasing use of MBIs in the perinatal period and their potential benefits for improving perinatal depression [23,24]. However, the evidence in this emerging area is still limited because the majority of research performed thus far has included small sample sizes and nonrandomized designs [21,25]. Moreover, few studies have been performed in LAMICs with most of these studies performed in high-income countries [25].

Over the last few years, internet-based MBIs and those accessed through apps have been increasingly considered for perinatal care. Online MBIs among nonpregnant women have shown small but significant effects on depression improvement with Hedges *g* of 0.29 [18]. However, the related literature for the pregnant population is sparse. Reported studies have been preliminary [26-29], aimed only at the postpartum period [30], treated mindfulness as a single component of a complex integrated intervention [31], or used only a pre-post test design [32]. Nevertheless, some pilot studies have supported online MBI as a promising technique to help expectant mothers reduce depressive symptoms [27,28].

More rigorous investigation of the effectiveness of online MBIs in the perinatal period has been lacking, and the need for a well-designed RCT to test the effectiveness in a larger sample and with longer-term follow up was identified. Thus, we carried out an 8-week self-guided smartphone-based mindfulness training intervention using an RCT with 32-week follow up and an attention control group from the early second trimester to the postpartum period among pregnant women deemed to be at risk of perinatal depression. The primary aim of this study was to evaluate the overall intervention effect on perinatal depression symptoms between the mindfulness training in pregnancy group and the attention control group. The effectiveness of mindfulness training in pregnancy on secondary outcomes (anxiety symptoms, perceived stress, positive and negative emotions, fatigue, sleep-related problems, memory, and fear of childbirth) were also explored. Further, we compared the effects of mindfulness training on depression remission at postintervention between the two groups. Finally, we tried to explore whether the intervention effects would differ between nulliparous and multiparous women.

Methods

Trial Design

This study was a single-center, two-parallel-armed, assessor-blinded, 1:1-allocated RCT with a 32-week follow up. Pregnant women who scored at or above the threshold for positive depressive symptoms on the Edinburgh Postnatal Depression Scale (EPDS) or Patient Health Questionnaire-9 (PHQ-9) were randomly assigned to the mindfulness intervention group (receiving 8-week smartphone-based mindfulness training) or attention control group (receiving 8-week regular WeChat health consultations). This study was approved by the ethical review board of the School of Nursing, Shandong University (2018-R-015). The trial was registered in the Chinese Clinical Trial Registry (ChiCTR1900028521) in December 2019 and no significant changes were made between the start of the trial and the registration confirmation.

Participants

The study was conducted between March 2018 and January 2020, and participants were recruited in the obstetrics clinic of a tertiary hospital in Jinan, Shandong, a city located in the east of China. The hospital provides perinatal services for around 5000 pregnant women each year.

Inclusion criteria for women to participate in the study included: (1) aged 18 years and over, (2) in the 12th to 20th week of gestation, (3) singleton pregnancy, (4) no plan to terminate pregnancy, (5) planned to receive antenatal examination and deliver in the study hospital, (6) completed junior high school education or above, (7) positive in depressive symptoms screening with an EPDS score>9 or a PHQ-9 score>4, (8) able to use the app on a smartphone for the study, and (9) able to understand and respond to the questionnaire.

Exclusion criteria were: (1) at risk of suicide or self-harm, (2) currently receiving psychiatric treatment or using psychiatric medications, (3) history of substance abuse or addiction in the past 6 months, (4) prior experience with mindfulness meditation, and (5) declined to participate.

Procedure

According to the national health policy in China, health status for all pregnant women is recorded in the 12th gestational week and they start prenatal visits regularly thereafter. Pregnant women were recruited from the records of the tertiary hospital where the study was based, during the time they are ordinarily required to report pregnancy-related information and receive preventive depression screening at their first regular visit in the obstetrics clinic. Informed consent for preventive psychological assessment was first obtained, and once consent was granted, printed questionnaires, including sociodemographics, pregnancy-related characteristics, and mental health indicators, were distributed for participants to complete and return. All participant records were screened according to the inclusion and exclusion criteria, and eligible participants were then contacted by a research assistant through telephone or WeChat (a popular instant social communicating software in China) within 2 weeks. Online informed consent for participation in the RCT was obtained through the online survey platform Wjx.cn. The RCT program was introduced to potential participants while discussing results of their preventive psychological assessment. This initial baseline evaluation of symptoms was classified as T1 for consented trial participants. The procedure and objectives of the trial were explained, and then online informed consent was obtained if they agreed to participate.

Recruited participants were randomly assigned to the mindfulness training during pregnancy or attention control group; the mindfulness group received 8-week mindfulness training and the control group received 8-week regular WeChat health consultation to control attention. Participants who were allocated to the mindfulness training group received the URL of the app through WeChat. They could use the app as long as they signed up. Data collection and assessment of outcomes took place over four time points in the follow-up period. T2 assessment took place at 4 weeks after allocation (intermediate period of intervention), T3 assessment took place at 8 weeks after allocation (endpoint of the intervention), T4 took place at 18 weeks after allocation (before childbirth), and T5 took place at 6 weeks after delivery. Follow-up assessments were collected by computer/smartphone-assisted self-administered surveys. All participants were awarded 2 yuan (US $0.30) when completing an assessment.

Intervention

Mindfulness Training During Pregnancy Group

The mindfulness training program was revised from Mindfulness Behavioral Cognitive Therapy (MBCT) developed by John Teasdale, Mark Williams, and Zindel Segal [33]. A psychologist with 5 years of mindfulness experience led the adaptation of the mindfulness training in pregnancy course. One obstetrician, one obstetric nurse, and two research assistants with mindfulness experience participated. Three principles guided the adaptation process: focus on perinatal depression and negative emotions, make physical adaptations for pregnant women, and simplify and shorten the practice properly, with each formal training limited to 25 minutes. The 8-week mindfulness training program contained eight sessions: week 1, *Understand mindfulness;* week 2, *Be in the present;* week 3, *Be mindful of negative emotions;* week 4, *Accept difficulties;* week 5, *Thoughts are just thoughts;* week 6, *Enjoy daily happiness;* week 7, *Mindful pregnancy and childbirth;* and week 8, *Continued mindfulness practice*. Each session was composed of thematic curriculum as well as formal and informal training lasting for 1 week. The thematic curriculum was provided through text, audio, and visual materials at the beginning of each session. Formal mindfulness training techniques were then introduced, and users were invited to continue practice, following the recordings and writing in the mindfulness journal, for 6 days per week. Formal mindfulness training included body scan, mindful breathing, mindful stretching, and mindful meditation lasting 15-25 minutes per day. Informal training was also recommended to be practiced every day, including pausing in the midst of daily life, mindful eating, mindful walking, and 3-minute breathing practices. Details on the components of the mindfulness training in pregnancy course are provided in Table 1.

Table 1. Content of mindfulness training during pregnancy intervention.

Week	Subject	Thematic curriculum	Formal training	Informal training
1	Understand mindfulness	1. Risks of negative emotions in pregnancy 2. What is mindfulness and how can it be useful for addressing symptoms 3. Raisin exercise	Body scan	Pausing in the midst of daily life
2	Be in the present	1. Being and Doing models 2. Be in the present 3. Unpleasant events journal	Mindful breathing	Mindful eating and mindful walking
3	Be mindful of negative emotions	1. Recognize the habitual stress response 2. List of negative thinking 3. 3-minute breathing space	Mindful breathing Mindful stretching	3-minute breathing space
4	Accept difficulties	1. Difficult communication journal 2. Learn to accept difficulties	Mindful meditation (be with difficulties)	3-minute breathing space
5	Thoughts are just thoughts	1. Recognize thoughts without judgment 2. Let thoughts be	Mindful meditation (be with thoughts)	3-minute breathing space
6	Enjoy daily happiness	1. Be mindful of happiness 2. Pleasant events journal	Body scan	Pausing in the midst of daily life
7	Mindful pregnancy and childbirth	1. Be mindful of emotions caused by pregnancy 2. Feel fetal movement mindfully	Body scan Mindful stretching (childbirth)	3-minute breathing space
8	Continued mindfulness practice	1. Discussion on awareness of emotions and stress responses 2. Consider continued mindfulness practice	Self-directed	Self-directed

The mindfulness training program was delivered through a custom-built mobile app called *Spirits Healing* in Chinese. It was available in both the Android and iOS operating systems in mainland China. The *Spirits Healing* app provided reading materials, recordings for guided practice, and videos. Participants were able to navigate contents and make notes in the app. The mindfulness training program automatically updated every day and participants practiced according to their own schedules. A message to remind participants to utilize the mindfulness training program was sent every week by WeChat. Participants were awarded 2 yuan (US $0.30) for completion of each week of training. The app was debugged three times during the trial due to adaptation of phone systems, but no changes related to intervention content were made. For safety and destigmatization, participants were reminded that this app is not equivalent to psychotherapy and were referred to professional support when necessary. Visual representations of the app content are shown in Figure 1. Additional details relating to the intervention construction process can be found in Multimedia Appendix 1.

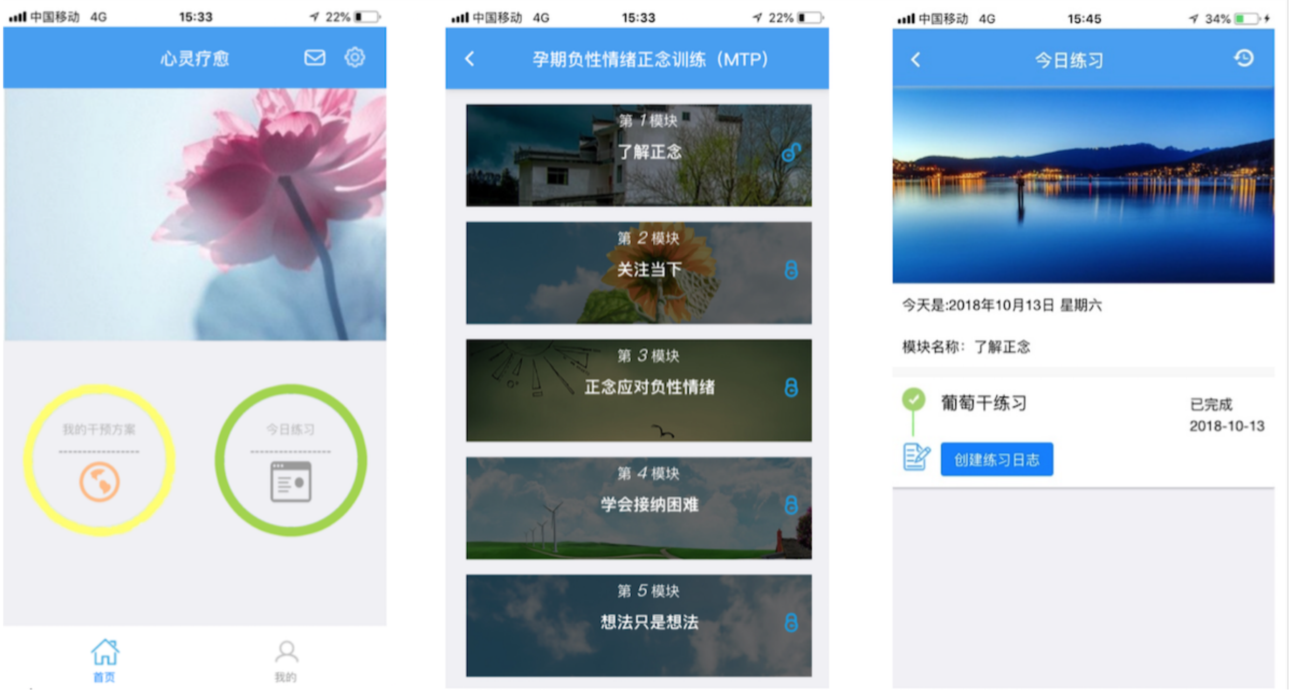


Figure 1. Screenshots of Spirits Healing mindfulness app.

今天是:2018年10月13日 星期六

模块名称：了解正念

Attention Control Group

Pregnant women allocated to the attention control group received 8-week regular WeChat health consultations as an attention control. Health consultations were provided by a clinically trained nursing assistant with experience in prenatal care using the WeChat app. The schedule for routine prenatal care was sent to participants at the time they were assigned to their group. The nursing assistant contacted participants in the attention control group once every week for 8 weeks to ask about recent health status using the following script: “Hello, Ms. X. How are you feeling this week?” The content of the consultations included discussion of recent medical examinations, outpatient appointments, and assistance with arrangements for inpatient care.

Measures

Depression symptoms as assessed by the EPDS was the primary outcome in this study. The EPDS is a self-report scale that assesses depressive symptoms experienced within the last week during both prenatal and postnatal periods [34]. The EPDS contains 10 items with responses on a 4-point Likert scale ranging from 0 to 3, where higher scores represent greater intensity of a depressive symptom. The EPDS was recommended as a valid depression screening tool across different cultures and different trimesters in a validation study review [35], and the cut-off score of 9/10 was used to identify positive depressive symptoms for screening purposes in this study. Additionally, another commonly used depression screening tool, the PHQ-9 developed on the basis of Diagnostic and Statistical Manual of Mental Disorders-IV criteria, was also used in the screening period with a cutoff of 4/5 indicative of positive depression criteria [36].

Secondary outcomes consisted of multidimensional health issues for perinatal women, including anxiety symptoms, perceived stress, positive affect, negative affect, sleep-related problems, fatigue, prospective memory, retrospective memory, and fear of childbirth. Anxiety in the previous 2 weeks was evaluated by the 7-item Generalized Anxiety Disorder scale, a clinically useful assessment for detection of symptoms of anxiety in the perinatal period [37]. We assessed perceived stress of pregnant women using the 4-item validated Perceived Stress Scale to measure the degree in the past month that their situation was appraised as uncontrollable, unpredictable, and overwhelming [38]. The Positive and Negative Affect Schedule [39] was used to measure individuals’ agreement and endorsement on statements related to positive and negative affect. Sleep-related problems in the past month were evaluated by the self-administered Pittsburgh Sleep Quality Index questionnaire [40] and the degree of fatigue in the past week was evaluated by the 9-item Fatigue Severity Scale [41]. Subjective prospective and retrospective memory failures in daily life were self-rated by participants using the Prospective and Retrospective Memory Questionnaire [42]. In addition, the Wijma Delivery Expectancy Questionnaire [43] with items related to women’s cognitive appraisal regarding the delivery process was used to measure the level of fear of childbirth. For all measurements other than positive affect, higher scores represent worse mental health outcomes, and for positive affect, higher scores represent a higher positive affect. Time points for the outcome indicators assessed are shown in Multimedia Appendix 2.

Sociodemographic characteristics such as age, gestational age, BMI, education level, work status, marital status, family economic status, and pregnancy-related characteristics were self-reported by the participants at baseline (T1). Study researchers were allowed to collect additional clinical data on participants from medical records following birth.

Sample Size

Previous meta-analyses indicated that mindfulness practice during pregnancy reduced depression scores with an effect size of 0.59 [44] and self-help mindfulness resulted in small to medium effect sizes on anxiety/depression [45]. On the basis of a medium intervention effect (Cohen *d*=0.50), an estimated sample size of 128 was required to compare between-group differences on depression at postintervention with 80% statistical power, a 1:1 allocation rate, and a two-tailed significance level of .05. In the above-mentioned meta-analysis of self-help mindfulness [45], on average, 27% of the participants were lost to follow up in the postintervention assessment. Considering a 30% attrition rate, a final sample of 168 with 84 individuals in each group was required. The sample size was calculated using G*Power [46].

Randomization and Blinding

The trial used a simple randomization approach. The random number sequence was generated by a researcher who did not participate in this study using the Random Number Generators function in SPSS 23.0. The sequence was kept in sealed, opaque, numbered envelopes. When each participant was enrolled, a research assistant opened the envelope in sequence and assigned the participant to the group. Participants received allocation according to the order of enrollment. Throughout the RCT procedure, the assessor who collected follow-up data did not know the assignment. Considering the use of the smartphone app only in the mindfulness intervention, the participants were able to infer the assignment.

Intervention Fidelity

Logs of practice on formal mindfulness training were recorded and used to evaluate the fidelity of mindfulness training. At least 3 days of practice per week was considered a completed training week, and at least 4/8 completed training weeks were considered as completion of the mindfulness training program. The completion rate in this study was calculated as the percentage of participants who completed the mindfulness training divided by the number of participants who received the intervention.

Statistical Analysis

Statistical analyses were performed with SPSS 23.0. The primary analysis used an intention-to-treat (ITT) approach. Normality of outcomes at baseline was visually examined, and negative affect, sleep, and positive affect were log-transformed due to non-normal data distributions. No more than 5 individuals for each measure at T1 had a single entry missing, which were considered as missing at random and were imputed with the median of the relevant item. Independent sample *t* tests and χ^2 tests were used to compare the baseline characteristics between the intervention and control groups, and between the study sample and dropout sample. To assess the population-averaged mindfulness training vs attention control intervention effect on outcomes, generalized estimating equations (GEEs) were formulated. This approach has been recommended because it is able to handle missing data appropriately and remains stable in different correlation matrices [47]. For GEEs, the participants and assessment time points were designated as subject variables and within-subject variables, respectively, with an exchangeable working correlation matrix and full maximum-likelihood estimation applied. The continuous outcome variables at five time points were the dependent variables. Main effects of group, time, and group × time interaction effects were examined. Testing of a simple main effect for group was also explored by examining differences between groups at each time point. To further assess the intervention effect on depression symptoms remission postintervention (T3), binary positive depressive symptoms at T3 and EPDS reductions from T1 to T3 were compared between mindfulness and attention control groups using logistic regression models, adjusting for the EPDS score at baseline and between-group imbalanced factor (intended pregnancy) after randomization. In addition, subgroup analysis was also conducted to explore the impact of parity (primipara/multipara) on the intervention effect.

Effect sizes are presented as Cohen *d* based on the ITT rule. Cohen *d* was calculated between groups and follow ups referring to baseline data. Effect sizes were considered to be small (*d*=0.2), medium (*d*=0.5), and large (*d*=0.8). A two-sided *P* value less than .05 in the primary analysis and less than .025 in the subgroup analysis after Bonferroni multiple-comparison correction were considered statistically significant.

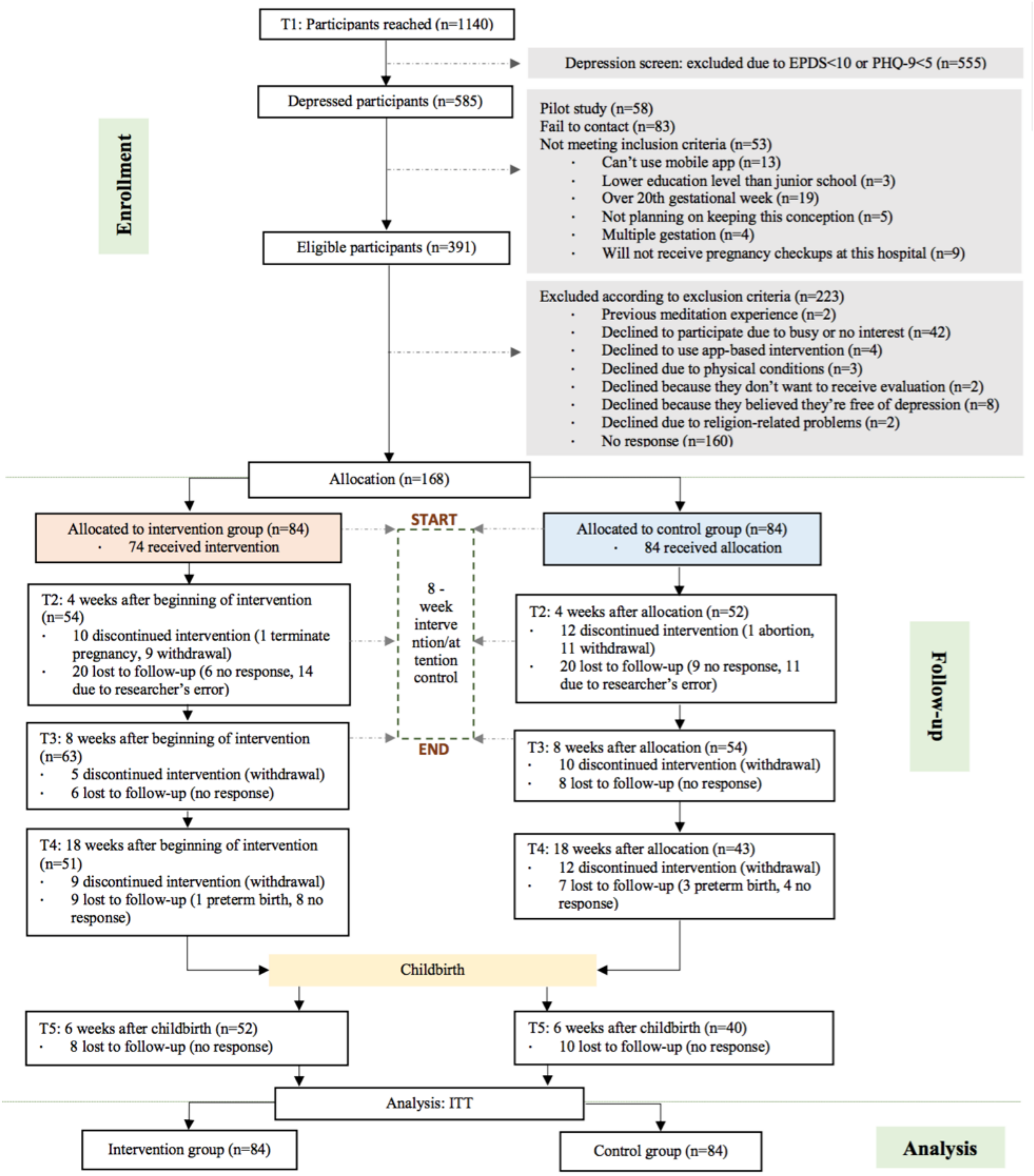
As suggested by Thabane [48], several sensitivity analyses were performed. First, the mindfulness training vs attention control intervention effect was analyzed according to per-protocol group and as-treated group comparisons in addition to the ITT group. Second, adjusted GEE models with a baseline imbalanced factor (intended pregnancy) were also established. Third, intervention effects were also evaluated in participants who completed different numbers of follow-up assessments. Details of the statistical analysis are provided in Multimedia Appendix 3.

Results

Recruitment and Participant Flow

Recruitment began in March 2018 and ended in June 2019, and follow-up assessments ended in January 2020. Figure 2 illustrates the full study flow diagram. A total of 1140 pregnant women were contacted and a final sample of 168 was allocated. During the whole follow-up period for the 168 participants, 24/84 (29%) participants dropped out in the mindfulness training group and 34/84 (41%) dropped out in the attention control group with the overall dropout rate reaching 34.5%. More than half of the participants completed at least 3 follow ups (94/168, 55.6%) and 56/168 (33.3%) of participants completed all follow ups. A logistic regression model with binary dropout status as the dependent variable showed that participants with an advanced gestational age at baseline tended to drop out more frequently during follow up (odds ratio [OR] 1.033, 95% CI 1.004-1.063; see Multimedia Appendix 4).

Figure 2. Participant flow chart. EPDS: Edinburgh Postnatal Depression Scale; PHQ-9: Patient Health Questionnaire-9; ITT: intention to treat.



Baseline Data

Table 2 shows the details of participant characteristics. Overall, all of the pregnant women were married, with an average age of around 30 years, and were at 99 gestational days (around 14 gestational weeks) at baseline. A total of 41.1% (69/168) of the participants screened positive in the EPDS and 92.3% (155/168) screened positive for depression symptoms in the PHQ-9. The mean gestational week was 16.8 (SD 1.068) at randomization. There was no significant difference between the mindfulness training and attention control groups at baseline except for intended pregnancy. More pregnant women planned this pregnancy in the mindfulness training group than in the attention control group (χ^2_1=8.4, *P*=.004).

Table 2. Participant characteristics at baseline.

Characteristics	Total (N=168)		ACG[a] (n=84)		MTPG[b] (n=84)		*P* value
	N	Value	N	Value	N	Value	
Age (years), mean (SD)	168	29.91 (4.015)	84	29.55 (4.21)	84	30.27 (3.80)	.24
Maternal age category, n (%)	168						.68
Not advanced (18-34 years)		140 (83.3)		71 (85)		69 (82)	
Advanced (>35 years)		28 (16.7)		13 (16)		15 (18)	
Gestational age (days), mean (SD)	168	98.81 (14.726)	84	100.85 (15.180)	84	96.77 (14.054)	.07
Weight gain, mean (SD)	156	1.77 (2.989)	77	1.98 (2.916)	79	1.57 (3.063)	.39
BMI before pregnancy, mean (SD)	162	21.81 (3.091)	79	21.44 (3.361)	83	22.16 (2.786)	.14
BMI now, mean (SD)	160	22.41 (2.935)	81	22.07 (2.994)	79	22.76 (2.850)	.14
Race, n (%)	168		84		84		.49
Han		167 (99.4)		83 (99)		84 (100)	
Hui		1 (0.60)		1 (1)		0 (0)	
Education years, mean (SD)	165	15.42 (2.361)	81	15.17 (2.323)	84	15.67 (2.386)	.18
Work status, n (%)	163		80		83		.22
Unemployed		38 (23.3)		22 (28)		16 (19)	
Employed		125 (76.7)		58 (72)		67 (81)	
Family monthly income per person (US $)[c], n (%)	152		74		78		.29
<309		1 (0.7)		1 (1)		0	
309-618		32 (21.1)		19 (26)		13 (17)	
618-926		46 (30.3)		19 (26)		27 (35)	
≥926		73 (48.0)		35 (47)		38 (49)	
Married, n (%)	164	164 (100)	81	81 (100)	83	83 (100)	
Parity, n (%)	168		84		84		.63
Primipara		109 (64.9)		53 (63)		56 (67)	
Multipara		59 (35.1)		31 (37)		28 (33)	
History of abortion, n (%)	168		84		84		.53
Yes		74 (44.0)		39 (46)		35 (42)	
No		94 (56.0)		45 (54)		49 (58)	
History of induced labor, n (%)	168		84		84		.50
Yes		9 (5.4)		3 (4)		6 (7)	
No		159 (94.6)		81 (96)		78 (93)	
History of embryo damage, n (%)	168		84		84		.17
Yes		33 (19.6)		13 (16)		20 (24)	
No		135 (80.4)		71 (84)		64 (76)	
Intended pregnancy, n (%)	162		80		82		.004
Yes		125 (77.2)		54 (68)		71 (87)	
No		37 (22.8)		26 (32)		11 (13)	
Severity of early pregnancy reaction[d], mean (SD)	150	2.70 (1.394)	71	2.62 (1.398)	79	2.77 (1.395)	.51
History of previous disease, n (%)	161		81		80		>.99
Yes		10 (6.2)		5 (6)		5 (6)	

Characteristics	Total (N=168)		ACG[a] (n=84)		MTPG[b] (n=84)		*P* value
	N	Value	N	Value	N	Value	
No		151 (93.8)		76 (94)		75 (94)	
EPDS[e], mean (SD)	168	8.27 (4.245)	84	8.55 (4.593)	84	7.99 (3.873)	.40
PHQ-9[f], mean (SD)	168	7.73 (3.295)	84	7.50 (3.036)	84	7.96 (3.538)	.36
GAD-7[g], mean (SD)	146	5.40 (3.239)	73	4.99 (3.204)	73	5.81 (3.243)	.13
PSS[h], mean (SD)	147	5.26 (2.315)	73	5.33 (2.267)	74	5.19 (2.374)	.72
PA[i], mean (SD)	143	27.29 (5.227)	71	26.85 (5.255)	72	27.72 (5.198)	.32
NA[j], mean (SD)	143	20.63 (5.734)	71	19.96 (5.111)	72	21.29 (6.254)	.17
PSQI[k], mean (SD)	139	6.85 (2.980)	72	6.58 (2.577)	67	7.13 (3.357)	.28
FSS[l], mean (SD)	145	43.16 (9.179)	72	43.78 (8.899)	73	42.55 (9.469)	.42
PM[m], mean (SD)	145	20.65 (6.017)	73	20.93 (6.106)	72	20.36 (5.954)	.57
RM[n], mean (SD)	145	19.54 (6.059)	73	19.90 (6.169)	72	19.17 (5.965)	.47
WDEQ[o], mean (SD)	138	45.54 (16.738)	68	46.26 (17.622)	70	44.84 (15.928)	.62

[a] ACG: attention control group.

[b] MTPG: mindfulness training during pregnancy group.

[c] Based on conversion of US $0.15=1 Chinese yuan at the time of writing.

[d] Participants rated the severity of early pregnancy reaction in the first 3 gestational months from 0 (the least serious) to 6 (the most serious).

[e] EPDS: Edinburgh Postnatal Depression Scale.

[f] PHQ-9: Patient Health Questionaire-9.

[g] GAD-7: Generalized Anxiety Disorder-7.

[h] PSS: Perceived Stress Scale.

[i] PA: Positive and Negative Affect Schedule-Positive Affect.

[j] NA: Positive and Negative Affect Schedule-Negative Affect.

[k] PSQI: Pittsburgh Sleep Quality Index.

[l] FSS: Fatigue Severity Scale.

[m] PM: Prospective and Retrospective Memory Questionnaire-Prospective Memory.

[n] RM: Prospective and Retrospective Memory Questionnaire-Retrospective Memory.

[o] WDEQ: Wijma Delivery Expectancy Questionnaire.

Fidelity

Of the 84 participants allocated to the mindfulness training group, 10 of them did not activate the app, which was considered refusing the allocation. All of the remaining participants received the mindfulness training program. On the basis of the ITT sample, the mean number of completed training weeks was 3 weeks (SD 2.701). As a whole, 44/84 participants completed at least 4 weeks of training and the total completion rate was 52.4%. In all, 7/84 (8%) participants completed the entire 8-week training program.

Overall Intervention Effect Between Groups

Primary Outcome: Depression

First, the overall test of the intervention effect on perinatal depression was performed through GEE analysis of the ITT sample. There was a significant time effect and group × time interaction effect on the change of the EPDS score (Table 3). As shown in Table 3 and Figure 3, the EPDS score at T2 decreased in both the mindfulness training and attention control groups. Thereafter, the EPDS score in the mindfulness training group continued to decline at T3, and remained at a low level at T4, but increased slightly postpartum. However, in the attention control group, the EPDS score increased markedly at T3 and then declined at T4 and T5. The mean difference between the two groups reached 2.82 points at T3. At T4, even though the between-group mean difference was not statistically significant, it was higher in the attention control group than in the mindfulness training group (Table 3). The between-group effect sizes for the EPDS were approximately medium at T3, as shown in Table 4. Analysis based on per-protocol and per-protocol–intervention complete samples showed the same results as the analysis from the ITT sample.

Table 3. Overall intervention effect on Edinburgh Postnatal Depression Scale scores using generalized estimated equation models.

Sample	Mean group difference (95% CI)	*P* value	Group effect		Time effect		Group × time effect	
			Wald chi-square (*df*=1)	*P* value	Wald chi-square (*df*=4)	*P* value	Wald chi-square (*df*=4)	*P* value
ITT[a]			1.4	.23	15.7	.003	16.2	.003
T1[b]	0.56 (–0.72 to 1.84)	.39						
T2[c]	0.14 (–1.65 to 1.94)	.88						
T3[d]	2.82 (0.93 to 4.71)	.003						
T4[e]	1.41 (–0.52 to 3.34)	.15						
T5[f]	–1.11 (–3.00 to 0.79)	.25						
PP[g]			0.4	.54	13.8	.008	16.0	.003
T1	0.04 (–1.22 to 1.30)	.95						
T2	–0.46 (–2.24 to 1.32)	.61						
T3	2.43 (0.58 to 4.28)	.01						
T4	1.08 (–0.84 to 2.99)	.27						
T5	–1.14 (–3.01 to 0.74)	.24						
PP-IC[h]			1.5	.23	12.6	.01	16.1	.003
T1	0.40 (–1.04 to 1.84)	.59						
T2	0.35 (–1.38 to 2.08)	.69						
T3	3.02 (1.17 to 4.86)	.001						
T4	1.28 (–0.92 to 3.49)	.23						
T5	–0.84 (–3.02 to 1.34)	.45						

[a] ITT: intention-to-treat; n=84 mindfulness training group, n=84 attention control group.

[b] T1: baseline assessment.

[c] T2: 4 weeks after group allocation.

[d] T3: 8 weeks after group allocation.

[e] T4: 18 weeks after group allocation.

[f] T5: 6 weeks after delivery.

[g] PP: per-protocol; n=74 mindfulness training group, n=94 attention control group.

[h] PP-IC: per-protocol intervention completed; n=44 mindfulness training group, n=84 attention control group.

Figure 3. Overall intervention effect on Edinburgh Postnatal Depression Scale (EPDS) scores using the generalized estimated equation model in (A) intention-to-treat (ITT), (B) per-protocol (PP), and (C) per-protocol–intervention complete (PP-IC) samples. *significant between-group mean difference. ACG: attention control group; MTPG: mindfulness training in pregnancy group.

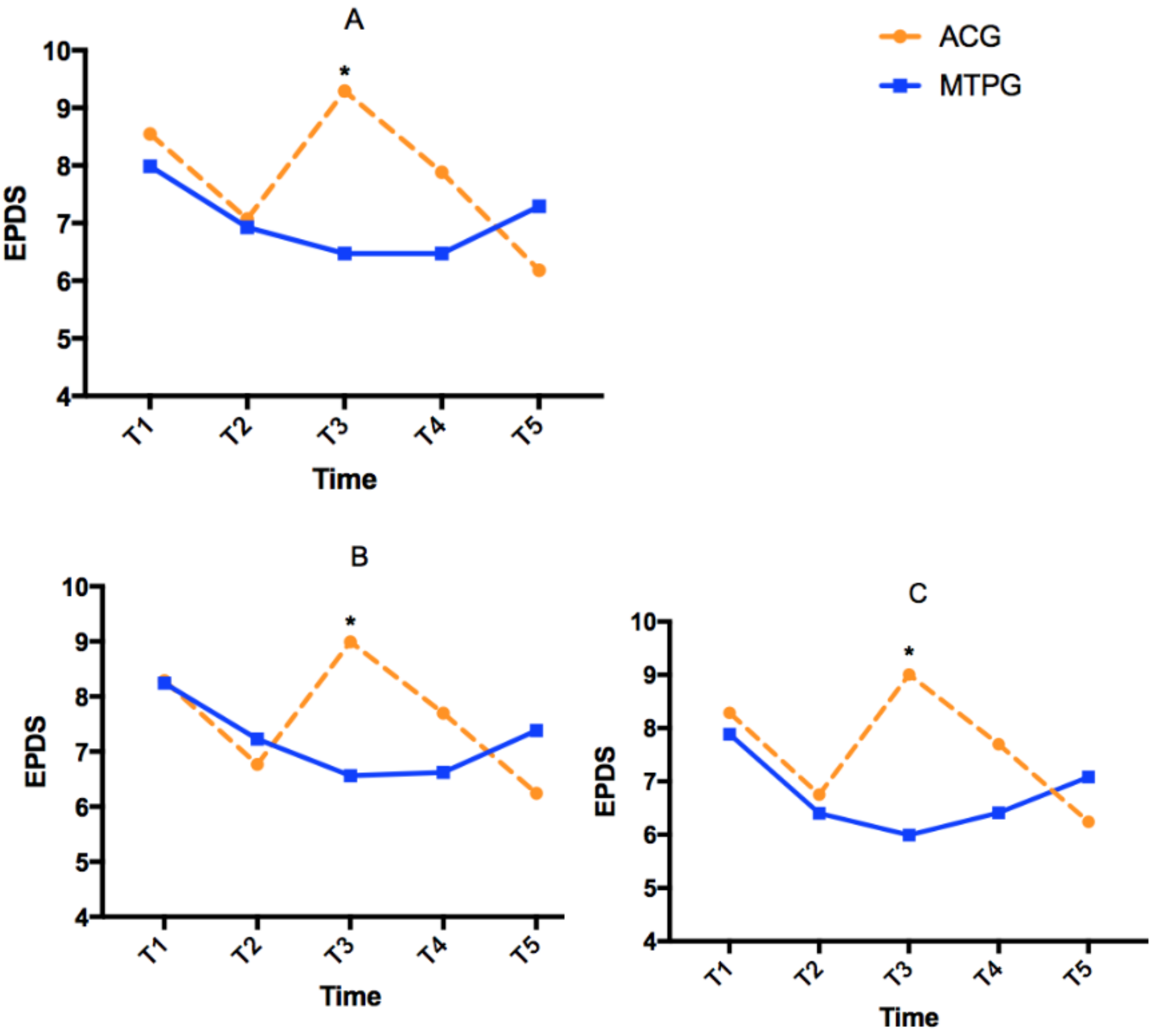


Table 4. Mean scores and effect sizes between and within groups for outcome measures.

Measure	ACG[a]			MTPG[b]			Between-group difference	
	Mean (SD)	Cohen *d*	*P* value	Mean (SD)	Cohen *d*	*P* value	Cohen *d*	95% CI
EPDS[c]								
T1[d]	8.55 (4.593)	reference	N/A[e]	7.99 (3.873)	reference	N/A	0.13	–0.17 to 0.43
T2[f]	7.63 (5.217)	0.19	.29	6.78 (4.816)	0.28	.11	0.17	–0.21 to 0.55
T3[g]	9.09 (6.241)	–0.10	.58	6.49 (4.497)	0.36	.03	0.47	0.10 to 0.84
T4[h]	8.02 (6.002)	0.10	.62	6.14 (4.552)	0.44	.01	0.35	–0.06 to 0.76
T5[i]	6.25 (5.098)	0.47	.01	6.77 (4.693)	0.28	.12	–0.11	–0.52 to 0.31
GAD-7[j]								
T1	4.99 (3.204)	reference	N/A	5.81 (3.243)	reference	N/A	–0.25	–0.58 to 0.07
T2	4.23 (2.975)	0.25	.18	4.50 (3.720)	0.38	.03	–0.08	–0.46 to 0.30
T3	5.56 (4.974)	–0.14	.46	4.46 (2.945)	0.44	.01	0.26	–0.10 to 0.63
T4	5.93 (4.079)	–0.26	.17	4.32 (2.773)	0.49	.01	0.46	0.04 to 0.87
T5	4.60 (3.967)	0.11	.58	4.32 (2.995)	0.48	.01	0.08	–0.33 to 0.49
PSS[k]								
T1	5.33 (2.267)	reference	N/A	5.19 (2.374)	reference	N/A	0.06	–0.26 to 0.38
T2	5.45 (2.403)	–0.05	.77	4.81 (2.782)	0.15	.41	0.25	–0.14 to 0.63
T3	6.09 (3.630)	–0.25	.18	5.22 (2.726)	–0.01	.94	0.27	–0.10 to 0.63
T4	6.09 (3.456)	–0.26	.20	5.24 (2.389)	–0.02	.92	0.29	–0.12 to 0.69
T5	5.95 (2.087)	–0.28	.16	5.90 (2.234)	–0.31	.09	0.02	–0.39 to 0.43
PA[l]								
T1	26.85 (5.255)	reference	N/A	27.72 (5.198)	reference	N/A	–0.17	–0.50 to 0.16
T3	24.20 (7.105)	0.42	.03	27.37 (5.857)	0.06	.71	–0.49	–0.87 to –0.12
T4	25.98 (7.220)	0.14	.50	28.26 (6.589)	–0.09	.62	–0.33	–0.75 to 0.08
T5	28.05 (7.887)	–0.18	.39	26.94 (6.150)	0.14	.45	0.16	–0.25 to 0.57
NA[m]								
T1	19.96 (5.111)	reference	N/A	21.29 (6.254)	reference	N/A	–0.23	–0.56 to 0.10
T3	18.59 (6.703)	0.23	.20	18.65 (6.479)	0.41	.02	–0.01	–0.38 to 0.36
T4	21.02 (7.333)	–0.17	.41	19.40 (5.852)	0.31	.09	0.24	–0.17 to 0.66
T5	19.60 (7.493)	0.06	.77	18.48 (5.043)	0.49	.01	0.18	–0.23 to 0.59
PSQI[n]								
T1	6.58 (2.577)	reference	N/A	7.13 (3.357)	reference	N/A	–0.18	–0.52 to 0.15
T3	6.19 (2.514)	0.15	.41	6.33 (3.653)	0.23	.20	–0.04	–0.42 to 0.33
T4	7.07 (2.640)	–0.19	.33	7.00 (3.156)	0.04	.83	0.02	–0.39 to 0.43
T5	7.59 (3.768)	–0.31	.14	8.12 (3.585)	–0.29	.13	–0.14	–0.56 to 0.27
FSS[o]								
T1	43.78 (8.899)	reference	N/A	42.55 (9.469)	reference	N/A	0.13	–0.19 to 0.46
T3	40.57 (11.812)	0.31	.09	40.75 (9.657)	0.19	.28	–0.02	–0.39 to 0.35
T4	43.21 (9.918)	0.06	.75	39.84 (9.678)	0.28	.12	0.34	–0.07 to 0.75
T5	38.50 (11.073)	0.53	.01	38.23 (11.758)	0.40	.03	0.02	–0.39 to 0.44

Measure	ACG[a]			MTPG[b]			Between-group difference	
	Mean (SD)	Cohen *d*	*P* value	Mean (SD)	Cohen *d*	*P* value	Cohen *d*	95% CI
PM[p]								
T1	20.93 (6.106)	reference	N/A	20.36 (5.954)	reference	N/A	0.09	–0.23 to 0.42
T3	20.34 (7.455)	0.09	.64	22.56 (6.995)	–0.34	.05	–0.31	–0.68 to 0.07
T5	21.25 (8.755)	–0.04	.84	23.04 (7.914)	–0.38	.04	–0.21	–0.63 to 0.20
RM[q]								
T1	19.90 (6.169)	reference	N/A	19.17 (5.965)	reference	N/A	0.12	–0.21 to 0.45
T3	19.58 (7.399)	0.05	.80	21.19 (7.238)	–0.30	.08	–0.22	–0.59 to 0.15
T5	21.00 (8.697)	–0.15	.48	22.40 (8.117)	–0.45	.02	–0.17	–0.58 to 0.25
WDEQ[r]								
T1	46.26 (17.622)	reference	N/A	44.84 (15.928)	reference	N/A	0.08	–0.25 to 0.42
T2	46.25 (19.312)	0.00	>.99	37.46 (21.055)	0.40	.03	0.44	0.05 to 0.82
T3	49.43 (18.283)	–0.18	.34	44.10 (20.121)	0.04	.81	0.28	–0.10 to 0.65
T4	48.42 (19.267)	–0.12	.55	46.04 (20.564)	–0.07	.73	0.12	–0.29 to 0.53

[a] ACG: attention control group.

[b] MTPG: mindfulness training during pregnancy group.

[c] EPDS: Edinburgh Postnatal Depression Scale.

[d] T1: baseline assessment.

[e] N/A: not applicable.

[f] T2: 4 weeks after group allocation.

[g] T3: 8 weeks after group allocation.

[h] T4: 18 weeks after group allocation.

[i] T5: 6 weeks after delivery.

[j] GAD-7: Generalized Anxiety Disorder-7.

[k] PSS: Perceived Stress Scale.

[l] PA: Positive and Negative Affect Schedule-Positive Affect.

[m] NA: Positive and Negative Affect Schedule-Negative Affect.

[n] PSQI: Pittsburgh Sleep Quality Index.

[o] FSS: Fatigue Severity Scale.

[p] PM: Prospective and Retrospective Memory Questionnaire-Prospective Memory.

[q] RM: Prospective and Retrospective Memory Questionnaire-Retrospective Memory.

[r] WDEQ: Wijma Delivery Expectancy Questionnaire.

Secondary Outcomes

Significant time and group × time interaction effects were found for anxiety in the ITT group (χ^2_4=18.8, *P*=.001; χ^2_4=13.1, *P*=.01) and per-protocol group (χ^2_4=19.6, *P*=.001; χ^2_4=13.3, *P*=.01) analyses, whereas only a significant time effect was found in the per-protocol–intervention complete group analysis (χ^2_4=21.7, P<.001). The anxiety score decreased at T2 and remained low in the mindfulness training group thereafter, but increased at T3 and T4 in the attention control group. A nearly medium between-group effect size was detected at T4 (Cohen *d*=0.46, 95% CI 0.04-0.87).

Significant group × time interaction effects were found for positive affect in the ITT (χ^2_4=8.4, *P*=.04), per-protocol (χ^2_4=8.2, *P*=.04), and per-protocol–intervention complete (χ^2_4=12.2 *P*=.007) analyses. The mindfulness training group maintained higher positive affect scores than the attention control group in the prenatal period, and reached a statistically significant mean difference at T3 for the ITT sample (–3.45, 95% CI –5.81 to –1.08, *P*=.004). The between-group effect size was medium (Cohen *d*=–0.49, 95% CI –0.87 to –0.12) at T3.

No significant between-group intervention effect was found on stress, log-transformed negative affect, log-transformed sleep, fatigue, log-transformed prospective memory, retrospective memory, and fear of childbirth. However, the between-group effect size on fear of childbirth was nearly medium (Cohen *d*=0.44, 95% CI 0.05-0.82) at T2. Details related to intervention effects on secondary outcomes are provided in Multimedia Appendix 5.

Postintervention Depression Symptoms Remission

For the binary EPDS variable at T3, fewer participants reported positive depressive symptoms (EPDS>9) in the mindfulness training group than in the attention control group (15/63, 24% vs 24/54, 44%; χ^2_1=5.6, *P*=.02). Results of the logistic regression model indicated that mindfulness training led to a 0.609-times reduction on the risk of positive antenatal depressive symptoms compared to the attention control (Table 5) at postintervention.

For EPDS reduction from T1 to T3, stepwise EPDS decreased scores ranging from 1 to 9 points were compared between groups; OR values remained statistically significant with EPDS score reduction>2. Mindfulness training showed a significant effect on EPDS score reduction over the attention control with the OR ranging from 3.471 to 27.986 (Table 5), indicating that mindfulness training had more potential to relieve depression symptoms than attention control.

Table 5. Intervention effect on depression symptoms remission at 8 weeks after allocation (T3) based on intention-to-treat analysis.

Depression symptom measure	ACG[a] (n=54), n (%)	MTPG[b] (n=63), n (%)	OR[c] (95% confidence limit) (reference=ACG)[d]
Positive depressive symptom (EPDS[e] >9) at T3			0.391 (0.164, 0.930)
No	30 (56)	48 (76)	
Yes	24 (44)	15 (24)	
EPDS decrease≥9 from T1[f] to T3			N/A[g]
Yes	0	7 (11)	
No	54 (100)	56 (89)	
EPDS decrease≥8 from T1 to T3			16.391 (1.507, 178.297)
Yes	1 (2)	10 (16)	
No	53 (98)	53 (84)	
EPDS decrease≥7 from T1 to T3			17.982 (1.798, 179.868)
Yes	1 (2)	12 (19)	
No	53 (98)	51 (81)	
EPDS decrease≥6 from T1 to T3			27.986 (2.907, 269.468)
Yes	1 (2)	15 (24)	
No	53 (98)	48 (76)	
EPDS decrease≥5 from T1 to T3			7.687 (1.980, 29.846)
Yes	5 (9)	18 (29)	
No	49 (91)	45 (71)	
EPDS decrease≥4 from T1 to T3			4.295 (14.79, 12.473)
Yes	9 (17)	24 (38)	
No	45 (83)	39 (62)	
EPDS decrease≥3 from T1 to T3			3.471 (1.275, 9.448)
Yes	14 (26)	26 (41)	
No	40 (74)	37 (59)	
EPDS decrease≥2 from T1 to T3			1.938 (0.838, 4.481)
Yes	20 (37)	31 (49)	
No	34 (63)	32 (51)	
EPDS decrease≥1 from T1 to T3			1.708 (0.752, 3.876)
Yes	24 (44)	35 (56)	
No	30 (56)	28 (44)	

[a] ACG: attention control group.

[b] MTPG: mindfulness training during pregnancy group.

[c] OR: odds ratio.

[d] calculated from logistic regression models with depression remission at T3 as the dependent variable, adjusting for EPDS score at baseline and intended pregnancy.

[e] EPDS: Edinburgh Postnatal Depression Scale.

[f] T1: baseline assessment.

[g] N/A: not applicable.

Subgroup Analysis

The mindfulness training vs attention control intervention effects on EPDS scores differed between a primipara and multipara state. In nulliparous women, the group × time interaction effect was statistically significant and the between-group mean difference reached 3.72 points on the EPDS at T3 (Table 6). However, no significant intervention effect was found in multiparous women (Table 6). Specifically, the EPDS score decreased at T3 in the mindfulness training group but increased in the attention control group in primipara, whereas it increased in both groups at T3 in multipara (Table 6 and Figure 4). However, the three-way interaction test (group × time × parity) did not reach statistical significance in the GEE model (χ^2_4=5.6, *P*=.24; see Multimedia Appendix 6). Results on subgroup analysis of anxiety and positive affect were consistent with those obtained for EPDS (see Multimedia Appendix 7).

Table 6. Overall intervention effect on Edinburgh Postnatal Depression Scale (EPDS) score by parity.

Parity status	Mean difference ACG[a] – MTPG[b] (95% CI)	*P* value	Group effect		Time effect		Group × time effect	
			Wald chi-square (*df*=1)	*P* value	Wald chi-square (*df*=4)	*P* value	Wald chi-square (*df*=4)	*P* value
Primipara (n=56 MTPG; n=53 ACG)			1.3	.26	10.2	.04	18.1	.001
T1[c]	0.72 (–0.81 to 2.25)	.36						
T2[d]	0.29 (–1.85 to 2.44)	.79						
T3[e]	3.72 (1.38-6.06)	.002						
T4[f]	1.03 (–1.36 to 3.42)	.40						
T5[g]	–1.32 (–3.60 to 0.96)	.26						
Multipara (n=28 MTPG; n=31 ACG)			0.2	.67	8.9	.06	2.4	.66
T1	0.30 (–1.98 to 2.59)	.79						
T2	–0.17 (–3.31 to 2.97)	.92						
T3	0.94 (–2.15 to 4.02)	.55						
T4	1.97 (–1.33 to 5.26)	.24						
T5	–0.67 (–4.05 to 2.72)	.70						

[a] ACG: attention control group.

[b] MTPG: mindfulness training during pregnancy group.

[c] T1: baseline assessment.

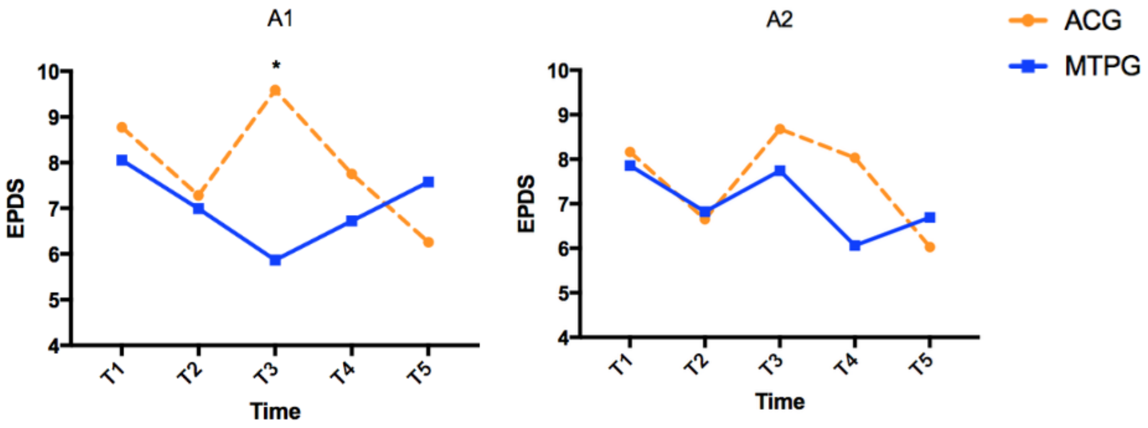
[d] T2: 4 weeks after group allocation.

[e] T3: 8 weeks after group allocation.

[f] T4: 18 weeks after group allocation.

[g] T5: 6 weeks after delivery.

Figure 4. Overall intervention effect on Edinburgh Postnatal Depression Scale (EPDS) scores by parity. A1: Model in primipara; A2: Model in multipara. *significant between-group mean difference.



Sensitivity Analysis

Sensitivity analysis in adjusted GEE models supported the primary results (Multimedia Appendix 8). In sensitivity analysis related to the number of follow ups (Multimedia Appendix 9), results from those who completed at least 2 or 3 follow ups showed a significant group × time interaction effect on the EPDS as well as in the primary analysis. However, mindfulness training vs attention control intervention effect trends were only found among participants who completed all follow ups. For anxiety, between-group mean differences at T3 reached statistical significance in participants who completed different numbers of follow ups in addition to the significant mindfulness training vs attention control intervention effect. Results from participants who completed at least 2 follow ups and all follow ups also supported the primary results on positive affect.

Discussion

Our findings support the overall intervention effect of smartphone-based mindfulness training for the reduction of prenatal depression and anxiety symptoms, and enhancement of positive affect. Smartphone-based mindfulness training led to a lower rate of positive depression screening at postintervention and a greater decline in depression symptoms from baseline to postintervention relative to the attention control group. In addition, exploratory analysis suggested that parity is a potential moderator for smartphone-based mindfulness training among pregnant women despite not reaching statistical significance. The results of this study provide the first robust evidence on the effectiveness of self-help, smartphone-based mindfulness training for perinatal depression symptomology using an RCT design, relatively large sample size, and 32-week follow up through the postpartum period.

The effect of the intervention on depression as measured by the EPDS was substantial. Pregnant women who received mindfulness training had at least a 2.471-times higher odds of a decreased EPDS score from baseline to postintervention compared with women in the attention control group. At postintervention assessment, women in the mindfulness group experienced a 60.9% reduction in risk for positive depressive symptoms, with a medium between-group effect size on the EPDS. Studies assessing associations between mindfulness training and prenatal depression have not indicated clear agreement in between-group postintervention effects [3,44,49]. This study supports the positive effect of mindfulness, and the resulting medium effect size on depression is comparable with the effects of in-person mindfulness interventions [21,25]. The effect size was higher than that reported in nonpregnant women (a small effect on depression, Hedges g=0.29) [18], revealing that smartphone-based mindfulness training is especially suitable for pregnant women, consistent with previous studies assessing pregnant women's preference for internet-based services [13,14]. Few studies of internet or smartphone-based mindfulness interventions in pregnancy have appeared in the literature, and those that are available mainly report preliminary effects on maternal depressive symptoms [27-29]. Thus, our study has provided the first robust evidence from an RCT in this area.

The sustained low levels of depressive symptoms for pregnant women who received mindfulness training are also encouraging. Even though no difference was found at the midpoint of the intervention, the significant between-group difference after the intervention was consistent with findings from previous in-person mindfulness studies during pregnancy [50-53]. This study found a longer-term effect of online mindfulness training in participants in contrast to other recent research that was more limited by the high rate of dropout [28]. Our results are also consistent with another study assessing depressive symptoms at each session of eight MBCT sessions, showing larger decreases of depression severity in the last phase of the intervention [54]. This suggests that the dose of an intervention is a promising factor potentially affecting the intervention effect. A further issue noted in this study is that the trend of a low level of depressive symptoms in the mindfulness group did not sustain throughout the postpartum stage. Previous studies of prenatal mindfulness interventions are also conflicting in this regard [30,53,55,56]. However, considering the growing audience for smartphone-based intervention formats, and the potential to reach larger audiences, the platform is quite advantageous. The treatment gap in the postpartum period may be further reduced through smartphone-based interventions, which can include reminders and functions to enhance frequency and compliance.

The exploration of additional benefits of mindfulness training in pregnancy provided promising results of a low anxiety level and high positive affect in follow ups. Mental health problems in pregnant women may be experienced as a cluster of symptoms during the perinatal period. In empirical research, multiple symptoms such as anxiety, stress, and fatigue were found to be correlated with perinatal depression [57-59]. Traditional perspectives on psychopathology presumed that some co-occurring symptoms originate from an underlying common cause [60,61]. Although determining the root causes of perinatal depression and secondary outcomes was outside the scope of this study, the findings of concurrent improvement on depression symptoms, anxiety symptoms, and positive affect utilizing mindfulness training nonetheless provide support for this approach. Consistent with other studies on mindfulness interventions [21,44,52], anxiety symptoms in this study were improved as well as depression symptoms. Another notable finding is the maintenance of positive affect in follow up, in contrast to less consistent findings on the effect of a mindfulness intervention on positive affect [21]. Duncan and Bardacke [19] reported a within-subject increase of positive affect in pregnant women through a pre-post test design trial. The between-group comparison and longer-term follow up in this study suggested that mindfulness training was helpful for pregnant women to maintain their initial positive affect, but did not constitute a pattern of increase. Mindfulness seems to particularly help pregnant women to maintain awareness of positive affect during the challenging perinatal period. Thus, mindfulness training in pregnancy may be a potentially cost-effective measure to resolve issues such as those included as secondary outcomes in this study. Cost-effectiveness in addressing multidimensional perinatal mental health is particularly appropriate for LAMICs. However, the full extent of costs involved in the intervention need to be specifically studied in future research.

Understanding the individual characteristics that may affect the efficacy of an intervention is key for evaluating suitability and usability. In this study, we tried to ascertain the moderating effect of parity. However, the nonsignificance of the three-way interaction test suggests that parity is only one potential factor influencing the intervention effect. Mindfulness interventions were previously reported to be less accessible with high loss to follow up among women who already had children [21]. Lack of time for participation and existing family commitments are important reasons for the likelihood that women may drop out of longitudinal research [62]. Reduction of depression symptoms among primipara who received mindfulness training was found only at postintervention in this study, pointing out a future direction for investigating the moderating effect over the shorter or longer term. Further studies are still needed for subgroups and vulnerable populations receiving this type of intervention.

This study expands on existing studies of mindfulness training in pregnancy in several key areas. First, it shows that prenatal mindfulness training can be extended through smartphone-based delivery, which may lead to reductions in demand on both therapists and service costs [63]. The cost savings would be especially helpful for LAMICs. Second, the demonstrated longer-term effect throughout pregnancy, and the multiple effects on depression, anxiety, and positive affect are beneficial to simplify intervention protocols. Third, the subgroup analysis on parity points to the importance for parous women to receive mental health services in the perinatal period. Additionally, the RCT design with attention control and sensitivity analysis reduced risk of bias and provided rigorous findings.

Several limitations should be noted. First, the potential impact of the dropout rate must be acknowledged. In this study, the overall rate of dropout was 34.5%, which is slightly higher than that reported for a recent in-person antenatal mindfulness program with a 28% overall dropout rate [20]. Difficulties with retention are a constant problem for self-help online research [28]. However, we also noted that the dropout rate at postintervention for this study was 22.0%, which is not larger than the estimated 30% attrition accounted for when calculating sample size, and did not limit the statistical power to test the interaction effect and immediate postintervention effect. The completion rate of the intervention is another limitation as it was unsatisfactory. Only 7 of 84 participants in the mindfulness group completed the entire 8-week training, and 44/84 participants completed the determined completion standard with a completion rate of 52.4%. Reported completion rates of interventions vary among online interventions, ranging from 21% to 74% [28,64] depending on definitions of completion rate and the intervention format. Even though the completion rate herein is within that range, misinterpretation of the intervention effect is a risk of bias from nonuse of the intervention. Third, 10 of 84 participants in the intervention group did not activate the mindfulness training app even after they agreed to participate in this program and were reminded by the research assistant. The high proportion of inactive participants in the intervention group weakens the strength of the RCT design. Fourth, the longer-term effect and the subgroup analysis related to parity were exploratory analyses in this study, and the sample sizes may be insufficient for comparing specific differences at T4, T5, and between subgroups. Future studies that focus on the longer-term effect or subgroups are suggested to calculate sample sizes based on the time point or group that had the lowest effect size. Participants in this study were able to infer their allocation and were unblinded. The multiple assessment of outcomes increases the risk for a type I error. In addition, the broad confidence intervals indicate the need for caution in interpreting the finding of depression remission at postintervention for generalizing the result. The sample included in this study was recruited in a single hospital, further limiting the generalizability of our findings. It is important for further researchers to conduct multicenter studies to examine these findings. Finally, even though China lags behind high-income countries in the widespread provision of perinatal mental health services, it is considered an upper-middle-income country and is in rapid development. The use of smartphones and networks in China may be much better than that in other LAMICs, especially low-income countries. More studies focusing on convenient and low-cost perinatal psychological interventions in these countries are still needed.

Acknowledgments

This study was funded by the Chinese National Funding of Social Sciences (grant number 17BSH054). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We also thank the China Scholarship Council for the support of YS while studying aboard. The authors would like to thank all doctors and nurses in obstetrics where the study was based, who provided enormous support in the recruitment and data collection of this study. The authors would also like to thank Dr Su at the Department of Biostatistics of Tulane University and Dr Cui at the School of Nursing of Shandong University who helped in resolving numerous statistical questions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Mindfulness training during pregnancy (MTP) program construction process.

[DOC File , 36 KB - jmir_v23i1e23410_app1.doc]

Multimedia Appendix 2
Timepoints of outcome indicators assessed.
[DOC File , 40 KB - jmir_v23i1e23410_app2.doc]

Multimedia Appendix 3
Statistical analysis.
[DOC File , 47 KB - jmir_v23i1e23410_app3.doc]

Multimedia Appendix 4
Analysis of dropouts.
[DOC File , 93 KB - jmir_v23i1e23410_app4.doc]

Multimedia Appendix 5
Overall intervention effects on secondary outcomes.
[DOC File , 2152 KB - jmir_v23i1e23410_app5.doc]

Multimedia Appendix 6
GEE model of subgroup analysis.
[DOC File , 52 KB - jmir_v23i1e23410_app6.doc]

Multimedia Appendix 7
Subgroup analysis on secondary outcomes.
[DOC File , 1554 KB - jmir_v23i1e23410_app7.doc]

Multimedia Appendix 8
Sensitivity analysis - adjusted model.
[DOC File , 2392 KB - jmir_v23i1e23410_app8.doc]

Multimedia Appendix 9
Sensitivity analysis for participants who completed different numbers of follow ups.
[DOC File , 2455 KB - jmir_v23i1e23410_app9.doc]

Multimedia Appendix 10
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 370 KB - jmir_v23i1e23410_app10.pdf]

References

1. Rallis S, Skouteris H, McCabe M, Milgrom J. A prospective examination of depression, anxiety and stress throughout pregnancy. Women Birth 2014 Dec;27(4):e36-e42. [doi: 10.1016/j.wombi.2014.08.002] [Medline: 25240846]
2. Pereira P, Lovisi G, Lima L, Legay L, de CSJ, Santos S. Depression during pregnancy: review of epidemiological and clinical aspects in developed and developing countries. In: Uehara T, editor. Psychiatric Disorders-Trends and Developments. Croatia: Intech; 2011:267.
3. Shi Z, MacBeth A. The Effectiveness of Mindfulness-Based Interventions on Maternal Perinatal Mental Health Outcomes: a Systematic Review. Mindfulness 2017;8(4):823-847 [FREE Full text] [doi: 10.1007/s12671-016-0673-y] [Medline: 28757900]
4. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. Lancet Psychiatry 2016 Oct;3(10):973-982 [FREE Full text] [doi: 10.1016/S2215-0366(16)30284-X] [Medline: 27650773]
5. Bauer AP, Knapp M, Lemmi V. The costs of perinatal mental health problems. London, UK: London School of Economics and Political Science; 2014. URL: https://www.centreformentalhealth.org.uk/sites/default/files/2018-09/costsofperinatal.pdf [accessed 2020-03-02] [WebCite Cache ID https://www.centreformentalhealth.org.uk/sites/default/files/2018-09/costsofperinatal.pdf]
6. Fisher J, Cabral de Mello M, Patel V, Rahman A, Tran T, Holton S, et al. Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review. Bull World Health Organ 2012 Feb 01;90(2):139G-149G [FREE Full text] [doi: 10.2471/BLT.11.091850] [Medline: 22423165]
7. Herrman H. Improving the mental health of women and girls: psychiatrists as partners for change. World Psychiatry 2016 Jun;15(2):190-191. [doi: 10.1002/wps.20316] [Medline: 27265719]

8. Dimidjian S, Goodman SH, Felder JN, Gallop R, Brown AP, Beck A. An open trial of mindfulness-based cognitive therapy for the prevention of perinatal depressive relapse/recurrence. Arch Womens Ment Health 2015 Feb;18(1):85-94. [doi: 10.1007/s00737-014-0468-x] [Medline: 25298253]
9. Guardino CM, Dunkel Schetter C, Bower JE, Lu MC, Smalley SL. Randomised controlled pilot trial of mindfulness training for stress reduction during pregnancy. Psychol Health 2014;29(3):334-349 [FREE Full text] [doi: 10.1080/08870446.2013.852670] [Medline: 24180264]
10. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
11. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
12. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
13. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
14. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the Use of Smartphone Apps for Health Information Among Pregnant Chinese Women: Mixed Methods Study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
15. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: Assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]
16. Kabat-Zinn J. Wherever you go, there you are: Mindfulness meditation in everyday life. New York: Hachette Books; 1994.
17. Kabat-Zinn J, Hanh T. Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness. London, UK: Bantam; 2009.
18. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
19. Duncan LG, Bardacke N. Mindfulness-Based Childbirth and Parenting Education: Promoting Family Mindfulness During the Perinatal Period. J Child Fam Stud 2010 Apr;19(2):190-202 [FREE Full text] [doi: 10.1007/s10826-009-9313-7] [Medline: 20339571]
20. Woolhouse H, Mercuri K, Judd F, Brown SJ. Antenatal mindfulness intervention to reduce depression, anxiety and stress: a pilot randomised controlled trial of the MindBabyBody program in an Australian tertiary maternity hospital. BMC Pregnancy Childbirth 2014 Oct 25;14(1):369 [FREE Full text] [doi: 10.1186/s12884-014-0369-z] [Medline: 25343848]
21. Matvienko-Sikar K, Lee L, Murphy G, Murphy L. The effects of mindfulness interventions on prenatal well-being: A systematic review. Psychol Health 2016 Dec;31(12):1415-1434. [doi: 10.1080/08870446.2016.1220557] [Medline: 27539908]
22. Donegan T. Mindfulness: An Intervention for Improving the Health of Mothers and Babies. Midwifery Today Int Midwife 2015(113):62-63. [Medline: 26309941]
23. Lavender TJ, Ebert L, Jones D. An evaluation of perinatal mental health interventions: An integrative literature review. Women Birth 2016 Oct;29(5):399-406. [doi: 10.1016/j.wombi.2016.04.004] [Medline: 27118000]
24. Hall HG, Beattie J, Lau R, East C, Anne Biro M. Mindfulness and perinatal mental health: A systematic review. Women Birth 2016 Feb;29(1):62-71. [doi: 10.1016/j.wombi.2015.08.006] [Medline: 26346905]
25. Van der Riet P, Francis L, Rees A. Exploring the impacts of mindfulness and yoga upon childbirth outcomes and maternal health: an integrative review. Scand J Caring Sci 2020 Sep;34(3):552-565. [doi: 10.1111/scs.12762] [Medline: 31797402]
26. Hulsbosch LP, Nyklíček I, Potharst ES, Meems M, Boekhorst MGBM, Pop VJM. Online mindfulness-based intervention for women with pregnancy distress: design of a randomized controlled trial. BMC Pregnancy Childbirth 2020 Mar 13;20(1):159 [FREE Full text] [doi: 10.1186/s12884-020-2843-0] [Medline: 32169030]
27. Yang M, Jia G, Sun S, Ye C, Zhang R, Yu X. Effects of an Online Mindfulness Intervention Focusing on Attention Monitoring and Acceptance in Pregnant Women: A Randomized Controlled Trial. J Midwifery Womens Health 2019 Jan 29;64(1):68-77. [doi: 10.1111/jmwh.12944] [Medline: 30695166]
28. Krusche A, Dymond M, Murphy SE, Crane C. Mindfulness for pregnancy: A randomised controlled study of online mindfulness during pregnancy. Midwifery 2018 Oct;65:51-57. [doi: 10.1016/j.midw.2018.07.005] [Medline: 30099285]
29. Matvienko-Sikar K, Dockray S. Effects of a novel positive psychological intervention on prenatal stress and well-being: A pilot randomised controlled trial. Women Birth 2017 Apr;30(2):e111-e118. [doi: 10.1016/j.wombi.2016.10.003] [Medline: 27810284]
30. Guo L, Zhang J, Mu L, Ye Z. Preventing Postpartum Depression With Mindful Self-Compassion Intervention: A Randomized Control Study. J Nerv Ment Dis 2020 Feb;208(2):101-107. [doi: 10.1097/NMD.0000000000001096] [Medline: 31868776]

31. Fonseca A, Monteiro F, Alves S, Gorayeb R, Canavarro MC. Be a Mom, a Web-Based Intervention to Prevent Postpartum Depression: The Enhancement of Self-Regulatory Skills and Its Association With Postpartum Depressive Symptoms. Front Psychol 2019;10:265. [doi: 10.3389/fpsyg.2019.00265] [Medline: 30873060]
32. Vieten C, Laraia BA, Kristeller J, Adler N, Coleman-Phox K, Bush NR, et al. The mindful moms training: development of a mindfulness-based intervention to reduce stress and overeating during pregnancy. BMC Pregnancy Childbirth 2018 Jun 01;18(1):201 [FREE Full text] [doi: 10.1186/s12884-018-1757-6] [Medline: 29859038]
33. Teasdale J, Williams J, Segal Z. The mindful way workbook: An 8-week program to free yourself from depression and emotional distress. New York: Guilford Publications; 2014.
34. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
35. Kozinszky Z, Dudas RB. Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. J Affect Disord 2015 May 01;176:95-105. [doi: 10.1016/j.jad.2015.01.044] [Medline: 25704562]
36. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
37. Simpson W, Glazer M, Michalski N, Steiner M, Frey BN. Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. Can J Psychiatry 2014 Aug;59(8):434-440 [FREE Full text] [doi: 10.1177/070674371405900806] [Medline: 25161068]
38. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
39. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
40. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
41. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989 Oct 01;46(10):1121-1123. [doi: 10.1001/archneur.1989.00520460115022] [Medline: 2803071]
42. Crawford JR, Smith G, Maylor EA, Della Sala S, Logie RH. The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample. Memory 2003 May;11(3):261-275. [doi: 10.1080/09658210244000027] [Medline: 12908675]
43. Wijma K, Wijma B, Zar M. Psychometric aspects of the W-DEQ; a new questionnaire for the measurement of fear of childbirth. J Psychosom Obstet Gynaecol 1998 Jun 07;19(2):84-97. [doi: 10.3109/01674829809048501] [Medline: 9638601]
44. Dhillon A, Sparkes E, Duarte RV. Mindfulness-Based Interventions During Pregnancy: a Systematic Review and Meta-analysis. Mindfulness 2017;8(6):1421-1437 [FREE Full text] [doi: 10.1007/s12671-017-0726-x] [Medline: 29201244]
45. Cavanagh K, Strauss C, Forder L, Jones F. Can mindfulness and acceptance be learnt by self-help?: a systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. Clin Psychol Rev 2014 Mar;34(2):118-129. [doi: 10.1016/j.cpr.2014.01.001] [Medline: 24487343]
46. Buchner AEE, Faul F, Lang AG. G*Power: Statistical Power Analyses for Windows and Mac. Allgemeine Psychologie und Arbeitpsychologie. URL: http://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html [accessed 2019-10-01]
47. Wang M. Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments. Adv Stat 2014 Dec 01;2014:1-11 [FREE Full text] [doi: 10.1155/2014/303728]
48. Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol 2013 Jul 16;13:92 [FREE Full text] [doi: 10.1186/1471-2288-13-92] [Medline: 23855337]
49. Lever Taylor B, Cavanagh K, Strauss C. The Effectiveness of Mindfulness-Based Interventions in the Perinatal Period: A Systematic Review and Meta-Analysis. PLoS One 2016;11(5):e0155720 [FREE Full text] [doi: 10.1371/journal.pone.0155720] [Medline: 27182732]
50. Shahtaheri E, Hosseinian S, Ebrahimi M, Jalilabadi Z, Mirderikvand F. The impact of stress reduction program based on mindfulness and conscious yoga on depression, perceived stress and quality of life in pregnant women. Acta Medica Mediterran 2016;32(5):2047-2050.
51. Pan W, Gau M, Lee T, Jou H, Liu C, Wen T. Mindfulness-based programme on the psychological health of pregnant women. Women Birth 2019 Feb;32(1):e102-e109. [doi: 10.1016/j.wombi.2018.04.018] [Medline: 29752225]
52. Yazdanimehr R, Omidi A, Sadat Z, Akbari H. The Effect of Mindfulness-integrated Cognitive Behavior Therapy on Depression and Anxiety among Pregnant Women: a Randomized Clinical Trial. J Caring Sci 2016 Sep;5(3):195-204 [FREE Full text] [doi: 10.15171/jcs.2016.021] [Medline: 27752485]

53. Dimidjian S, Goodman SH, Felder JN, Gallop R, Brown AP, Beck A. Staying well during pregnancy and the postpartum: A pilot randomized trial of mindfulness-based cognitive therapy for the prevention of depressive relapse/recurrence. J Consult Clin Psychol 2016 Feb;84(2):134-145 [FREE Full text] [doi: 10.1037/ccp0000068] [Medline: 26654212]
54. Felder JN, Segal Z, Beck A, Sherwood NE, Goodman SH, Boggs J, et al. An Open Trial of Web-Based Mindfulness-Based Cognitive Therapy for Perinatal Women at Risk for Depressive Relapse. Cogn Behav Pract 2017 Feb;24(1):26-37. [doi: 10.1016/j.cbpra.2016.02.002]
55. Vieten C, Astin J. Effects of a mindfulness-based intervention during pregnancy on prenatal stress and mood: results of a pilot study. Arch Womens Ment Health 2008;11(1):67-74. [doi: 10.1007/s00737-008-0214-3] [Medline: 18317710]
56. Bowen A, Baetz M, Schwartz L, Balbuena L, Muhajarine N. Antenatal group therapy improves worry and depression symptoms. Isr J Psychiatry Relat Sci 2014;51(3):226-231 [FREE Full text] [Medline: 25618288]
57. Paul S, Corwin EJ. Identifying clusters from multidimensional symptom trajectories in postpartum women. Res Nurs Health 2019 Apr 02;42(2):119-127. [doi: 10.1002/nur.21935] [Medline: 30710373]
58. Vismara L, Sechi C, Neri M, Paoletti A, Lucarelli L. Maternal perinatal depression, anxiety, fear of birth, and perception of infants' negative affectivity at three months. J Reprod Infant Psychol 2020 Nov 11:1-12. [doi: 10.1080/02646838.2020.1843612] [Medline: 33172285]
59. Pampaka D, Papatheodorou SI, AlSeaidan M, Al Wotayan R, Wright RJ, Buring JE, et al. Depressive symptoms and comorbid problems in pregnancy - results from a population based study. J Psychosom Res 2018 Sep;112:53-58. [doi: 10.1016/j.jpsychores.2018.06.011] [Medline: 30097136]
60. Borsboom D. Psychometric perspectives on diagnostic systems. J Clin Psychol 2008 Sep;64(9):1089-1108. [doi: 10.1002/jclp.20503] [Medline: 18683856]
61. Schmittmann VD, Cramer AO, Waldorp LJ, Epskamp S, Kievit RA, Borsboom D. Deconstructing the construct: A network perspective on psychological phenomena. New Ideas Psych 2013 Apr;31(1):43-53. [doi: 10.1016/j.newideapsych.2011.02.007]
62. Barnett J, Aguilar S, Brittner M, Bonuck K. Recruiting and retaining low-income, multi-ethnic women into randomized controlled trials: successful strategies and staffing. Contemp Clin Trials 2012 Sep;33(5):925-932 [FREE Full text] [doi: 10.1016/j.cct.2012.06.005] [Medline: 22732312]
63. Boggs JM, Beck A, Felder JN, Dimidjian S, Metcalf CA, Segal ZV. Web-based intervention in mindfulness meditation for reducing residual depressive symptoms and relapse prophylaxis: a qualitative study. J Med Internet Res 2014 Mar 24;16(3):e87 [FREE Full text] [doi: 10.2196/jmir.3129] [Medline: 24662625]
64. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]

Abbreviations

EPDS: Edinburgh Postnatal Depression Scale
GEE: generalized estimating equations
ITT: intention to treat
LAMIC: low- and middle-income country
MBCT: Mindfulness Behavioral Cognitive Therapy
MBI: mindfulness-based intervention
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
T1: baseline
T2: 4 weeks after allocation
T3: 8 weeks after allocation
T4: 18 weeks after allocation
T5: 6 weeks after delivery

Edited by G Eysenbach; submitted 11.08.20; peer-reviewed by R Eckhoff, M Lozano-Lozano, S Mukherjee; comments to author 30.09.20; revised version received 17.12.20; accepted 29.12.20; published 27.01.21.

Please cite as:
Sun Y, Li Y, Wang J, Chen Q, Bazzano AN, Cao F
Effectiveness of Smartphone-Based Mindfulness Training on Maternal Perinatal Depression: Randomized Controlled Trial
J Med Internet Res 2021;23(1):e23410
URL: http://www.jmir.org/2021/1/e23410/
doi:10.2196/23410
PMID:33502326

Original Paper

General Behavioral Engagement and Changes in Clinical and Cognitive Outcomes of Patients with Type 2 Diabetes Using the Time2Focus Mobile App for Diabetes Education: Pilot Evaluation

Bryan C Batch[1*], MD, MHS; Susan E Spratt[1*], MD; Dan V Blalock[2,3*], PhD; Chad Benditz[4*]; Andi Weiss[4*], MPH; Rowena J Dolor[5*], MD; Alex H Cho[5*], MD

[1]Division of Endocrinology, Metabolism and Nutrition, Duke University School of Medicine, Durham, NC, United States
[2]Center for Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Medical Center, Durham, NC, United States
[3]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, United States
[4]MicroMass Communications, Inc, Cary, NC, United States
[5]Division of General Internal Medicine, Department of Medicine, Duke University School of Medicine, Durham, NC, United States
[*]all authors contributed equally

Corresponding Author:
Bryan C Batch, MD, MHS
Division of Endocrinology, Metabolism and Nutrition
Duke University School of Medicine
DUMC 3031
Durham, NC, 27710
United States
Phone: 1 919 684 4005
Fax: 1 919 681 7796
Email: bryan.batch@duke.edu

Abstract

Background: Type 2 diabetes affects 30 million Americans, representing a significant cause of morbidity and mortality. Self-management support is an important component of chronic illness care and is a key pillar of the chronic care model. Face-to-face teaching and patient education materials suffer from being static or incompatible with mobile lifestyles. Digital apps provide a self-management support alternative that is convenient and scalable.

Objective: This pilot study tested the real-world deployment of a self-guided mobile app for diabetes education (Time2Focus app; MicroMass Communications Inc, Cary, NC), which utilizes evidence-based content and gamification to deliver an interactive learning experience.

Methods: Primary care providers were approached for permission to invite their patients to participate. Eligible patients were 18 to 89 years of age, had a diagnosis of type 2 diabetes, hemoglobin A1c (HbA1c) ≥8% and <12% in the past 3 months, an active online patient portal account (tied to the electronic health record), and access to an iOS or Android smartphone. Interested patients were emailed a baseline survey, and once this was completed, were sent instructions for downloading the Time2Focus app. After completing all 12 levels, participants were sent a follow-up survey. The primary outcome was the change in HbA1c. Secondary outcomes included medication adherence, self-care activities, self-reporting of physical activities, diabetes self-efficacy, illness perceptions, diabetes distress scale, and users' engagement with and rating of the app.

Results: Of 1355 potentially eligible patients screened, 201 were consented. Of these 201 patients, 101 (50.2%) did not download the app. Of the 100 participants (49.8%) who downloaded the app, 16 (16.0%) completed 0 levels, 26 (26.0%) completed 1 to 4 levels, 10 (10.0%) completed 5 to 11 levels, and 48 (48.0%) completed all 12 levels of the app and the follow-up survey. Those completing one or more levels had a mean pre/post-HbA1c change of –0.41% (compared to –0.32% among those who completed zero levels); however, the unadjusted two-tailed *t* test indicated no significant difference between the two groups (P=.73). Diabetes self-efficacy showed a large and significant increase during app usage for completers (mean change 1.28, P<.001, d=.83). Severity of illness perceptions showed a small but significant decrease during app usage for completers (mean change –0.51, P=.004, d=.43). Diabetes distress showed a small but significant decrease during app usage for completers (mean change –0.45, P=.006, d=.41). The net promoter score was 62.5, indicating that those who completed all levels of the app rated it highly and would recommend it to others.

Conclusions: Participants who engaged in all 12 levels of the Time2Focus mobile app showed an improvement in diabetes self-efficacy and a decrease in severity of illness perceptions. The decrease in HbA1c observed in app users relative to nonusers during this limited pilot study was not statistically significant. However, uptake and application of lessons learned from self-management support may be delayed. Further research is needed to address how to increase engagement through self-management support and to investigate if follow up over a longer period demonstrates a significant change in outcomes such as HbA1c.

(J Med Internet Res 2021;23(1):e17537) doi:10.2196/17537

KEYWORDS

mobile technology; diabetes; self management support; self efficacy; illness perception

Introduction

Diabetes affects 30 million Americans, and is a significant cause of morbidity and mortality. The care of diabetes and complications associated with the condition contribute to extraordinary expenditures each year, with the cost of diagnosed diabetes management in the United States reaching 327 billion dollars in 2017 [1].

Diabetes self-management support (SMS) provides patients with the knowledge and skills for implementing and sustaining the coping skills and behaviors needed to self-manage their diabetes on an ongoing basis [2]. SMS is an essential component of diabetes management, considering that diabetes is a condition in which the outcomes are heavily influenced by patient behaviors such as change in diet and physical activity. Further, SMS is an important component of chronic illness care and is a key pillar of the chronic care model, which is one of the frameworks upon which the patient-centered medical home concept is based. Prior research has shown that SMS is cost-effective, and is associated with improved patient knowledge and self-care behaviors, lower hemoglobin A1c (HbA1c), lower self-reported weight, reduced hospital admissions, reduced health care costs, and reduced risk of all-cause mortality [3,4]. However, evidence shows that only 5% to 7% of people eligible for SMS ultimately receive this support [4].

The traditional model of SMS training consists of in-person individual or group sessions. However, there are numerous challenges to the consistent and ongoing provision of such support. Traditional models have the disadvantage of being inherently limited by the need for patients to be face to face with a clinician or to attend a scheduled session at a fixed time and place. In addition, providers may not be skilled in communication techniques that have been shown to be effective in influencing behavior change.

The widespread availability of mobile smartphones and digital apps offers an alternative method of providing SMS to patients. Well-designed apps can incorporate the principles of evidence-based support for health behavior change. This method has the advantage of being convenient, patient-centered, economical, and scalable.

Despite the multitude of clinical trials focused on implementation of phone apps technology for SMS training, there continues to be a gap in the uptake of SMS training and mastery of training by patients who need it most. Further, studies of effectiveness of prior apps that include educational modules in improving clinical outcomes such as HbA1c are limited. A systematic review by Adu et al [5] explored considerations when designing and implementing apps for diabetes SMS. They noted that future designs of mobile apps for SMS need to include key elements of SMS education [5].

The Time2Focus digital app for diabetes education was developed by MicroMass Communications, Inc (Cary, NC), a firm specializing in patient and health care provider behavior change. The app is intended to help patients with type 2 diabetes build practical problem-solving skills through an interactive 12-level learning experience that uses gamification and evidence-based change support principles. Examples of topics covered include healthy eating, physical activity, and diabetes self-monitoring.

Here, we present the results of a pilot study conducted at Duke Health that tested a real-world deployment of Time2Focus, a self-guided mobile app for diabetes education. Outcomes assessed include HbA1c and change in self-reported health behaviors. All activities described below were approved by the Duke University Institutional Review Board.

Methods

Study Population, Recruitment, and Enrollment

Enrollment began in July 2017 and was completed in March 2018. Participants were eligible if they were ≥18 years of age, had a diagnosis of type 2 diabetes, HbA1c ≥8% (and <12%) within the past 3 months, an active MyChart patient portal account, and an iOS or Android smartphone.

Participants were excluded if they had a diagnosis of type 1 diabetes, were unable to provide consent, were legally blind, could not read, or did not read and understand English. The principal investigator of the study (AC) reached out to primary care providers at Duke Primary Care, a health system–affiliated network of primary care practices in central North Carolina. The network included 26 adult continuity practices at the time that recruitment for this study began. The principal investigator requested permission from providers to contact patients who were potentially eligible for the study. Participants were also identified via endocrinology clinic–based provider outreach. Recruitment letters were sent to these patients and signed by their primary care providers. The letters briefly described the study and allowed patients to opt out of being contacted. After informed consent was obtained, participants were contacted by MicroMass to introduce them to the Time2Focus mobile app.

Initial contacts were via email, and patients were asked to complete a baseline survey before they were sent instructions on how to download the app. Participants were instructed on the use of the app, and they were provided technical support as needed while they went through the 12 levels. The treatment group included all participants who successfully downloaded the app (n=100) and were eligible to begin interacting with it (regardless of completion of any level). Participants who consented to participate but did not download the app or complete any modules (n=101) comprised the control group for this analysis. Defining the control group in this way capitalized on the fact that both groups would have a similar willingness and ability to participate as the intervention group. Participant attrition that led to the final numbers in each cohort are discussed in further detail below.

Intervention

The goal of the Time2Focus app was to improve HbA1c, and patient confidence and skills in effectively managing diabetes by providing actionable, real-world guidance for diet, exercise, and glucose monitoring. The app was empirically based and incorporated components from evidence-based behavioral techniques, including goal-setting, problem-solving, feedback and reinforcement, and motivational interviewing. The app also leverages gamification principles to drive patients' self-management behaviors, including incentivizing participants' progress through positive feedback (positive messages when successfully completing a task), gradually adding complexity, and rewarding success (points earned for each module completed). Points are assigned for each task completed and accumulate over time. The total number of points through the course of use of the app served as a reminder to the participant regarding how far they progressed. There were also multiple challenges built into the app in the form of games.

The experience guided patients through progressive skill-building activities related to real-world situations. Unlike current mobile apps for type 2 diabetes, Time2Focus goes beyond simple tracking and patient education. The app was designed with the aim of improving patients' confidence (ie, self-efficacy) in their ability to carry out tasks, build problem-solving skills, and enhance goal-setting. Points earned and feedback given served to incentivize the key skills known to help support patient skill-building.

Time2Focus consists of 12 levels designed to enhance patients' skills and motivation with respect to several aspects of managing type 2 diabetes (eg, healthy eating, physical activity, self-monitoring blood glucose) (see Table 1).

The Time2Focus app was designed to have participants complete one level each week for a total of 12 weeks. Each level, on average, was designed to take about 1 hour to complete. Participants could start and stop each level as they wished, thereby personalizing their experience in using the app. Personalization was a central component in the app design because offering choice in the way one interacts with an intervention has been shown to promote autonomy, which in turn enhances intrinsic motivation for use of digital health interventions and engagement with behavior change [6]. Each level contained an assortment of articles, videos, challenges, and tracking tools. Once a participant completed the necessary requirements to finish each level, they received a text and email message to let them know when the next level would unlock and be available for them to start.

As part of the framework, Time2Focus encourages participants to build their skills through challenges. Participants had to attain a minimum score of 7 out of 10 on all challenges to progress to the next level. Participants who scored less than 70% could retake the challenges until they received a passing score.

Once the participants finished all 12 levels, they were sent a follow-up survey to complete. The follow-up survey included the same questions as found in the baseline survey with the addition of questions to assess the Time2Focus program. After completion of the follow-up survey, the participants received a congratulatory email, a US $25 gift card for compensation for completing each survey (totaling up to US $50), and entry to a raffle to win one of five iPad Mini tablets (ie, a participant who completed all 12 levels earned 12 entries into the raffle). Each participant was aware at the time they signed consent that they would receive a gift card and could possibly win an iPad Mini. The final gift was designed to be an incentive for completing the study.

Table 1. Summary of level learning objectives.

Level	Learning objectives
Level 1: Time2Focus Basics	Explain type 2 diabetes and the roles of insulin and blood glucose in the body
Level 2: Focus on Carbs	Distinguish carbohydrates from other types of food
Level 3: Focus on Physical Activity	Explain the effect of physical activity on blood glucose in type 2 diabetes
Level 4: Focus on Monitoring	Learn how to self-monitor and track blood glucose
Level 5: Mastery Challenges 1	Assess knowledge and comprehension of key concepts, including type 2 diabetes, insulin, and blood glucose
Level 6: Focus on Carb Planning	Identify targets for total carbohydrates for meals and snacks (based on carbohydrate counting)
Level 7: Focus on Making Choices	Identify protein foods, including carbohydrate proteins, and serving sizes
Level 8: Focus on Day-to-Day Decisions	Explain strategies for planning ahead for carbohydrate decisions, physical activity, and self-monitoring
Level 9: Making Decisions Away From Home	Identify strategies for planning and making decisions away from home
Level 10: Mastery Challenges 2	Assess knowledge and comprehension of key information and concepts, including nutrition, physical activity, and self-monitoring
Level 11: Keep Your Momentum Going	Establish habits, encourage decision-making, and facilitate goal-setting in managing type 2 diabetes
Level 12: Focus on the Future	Provide motivation and encouragement to patients that type 2 diabetes can be managed

Demographic Variables

Demographic and other baseline variables collected included race, sex, age, weight, BMI, and hypertension as recorded in the electronic health record (EHR).

Engagement

The degree of engagement was defined a priori based on the number of levels of the app that the participant completed. Levels of engagement were broken down into the following groups: participants who completed 0 levels, 1 to 4 levels, 5 to 11 levels, and 12 levels. Level 5 was the first mastery level to test what the participants had learned and to demonstrate their ability to apply these new skills. This level consisted of challenges only, and did not include articles or videos. Completing levels 5 and above indicated a markedly increased level of engagement. Thus, we a priori dichotomized engagement as low (downloading the app and completing anywhere from 0 to 4 levels) or high (completing level 5 and above).

Primary Outcome

The primary outcome was a change in HbA1c after participants concluded their use of the app. HbA1c data were abstracted from the EHR, utilizing naturally occurring measurements obtained in the normal course of clinical care.

Secondary Outcomes: Change in Self-Reported Health Behaviors and Perceptions of the App

Each study participant was required to complete the baseline survey before downloading the Time2Focus app and starting the program. The survey took about 10 minutes to complete, which gathered information regarding the following secondary outcomes: medication adherence, self-care activities, self-report of physical activity, diabetes self-efficacy, illness perceptions, diabetes distress scale, and users' engagement with and rating of the app.

The Voils scale (scores from 1 to 5) was used to measure medication adherence [7]; a higher score indicates that an individual is less adherent to medications. The Stanford scale was used to measure diabetes self-efficacy [8]; a higher score (scale from 1 to 10) denotes that an individual possesses more self-efficacy to manage their diabetes. Illness perceptions were measured using the Brief Illness Perception Questionnaire [9]; a higher score (scale from 0 to 10) denotes more severe perceptions about one's illness and has been shown to be associated with lowered perceptions of one's ability to cope with illness. Diabetes distress was measured using 4 questions from the Diabetes Distress Scale (scale from 0 to 5) [10]; a higher score communicates that an individual has more diabetes-related stress. The baseline survey also included questions about self-care activities (eg, how often they engaged in self-care per week, and how often on average they checked their blood sugar) and self-reported physical activity (a higher score means more physical activity reported).

Analysis

Independent-sample *t* tests and χ^2 tests were used to examine differences in baseline characteristics for continuous and dichotomous variables, respectively, across treatment and control groups as well as between participants with low and high levels of engagement. To assess change over time in HbA1c and self-reported health behaviors, regression analyses were used, controlling for baseline values of each outcome of interest. To facilitate interpretation, the dichotomous treatment variable in all regression analyses was dummy-coded such that "0" represented the control group and "1" represented the treatment group. In all regression analyses, including those with interaction terms to test the moderation of effects, all dichotomous and categorical covariates (with the exception of treatment) were coded effects such that the regression weights represent a comparison of each group with the average across all groups. For different analyses, mean clinical values, changes in mean clinical values, standardized β weights, or Cohen *d* effect size, estimates are presented to maximize interpretation. All analyses were performed in R version 3.5.3 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Study Population, Recruitment, and Enrollment

Seventy-eight primary care providers representing 23 different adult continuity practices gave permission for their patients to be contacted about this study. Of the 1355 potentially eligible patients screened, 780 were ineligible, 169 declined, 205 could not be reached, and 201 were consented. Among the 201 consented participants, 143 (71.1%) completed the baseline survey and 100 (49.8%) downloaded the Time2Focus app, who were eligible to start the 12-week Time2Focus program.

Baseline Demographics and Survey Results

The baseline characteristics of the participants are summarized in Table 2. The participants on average were 57.3 years old, 54.7% (110/201) were white, and there was an approximately equal sex ratio. They had a mean BMI of 35.8 and HbA1c of 9.02%. Among the 201 participants, over half had hypertension and approximately 10% had hyperlipidemia. These characteristics were also compared across app downloaders (100/201, 49.8%) and nondownloaders (101/201, 50.2%). Nonwhite participants were significantly less likely to download the app than white participants (χ^2=4.25, *P*=.04); however, no other baseline characteristics were significantly different across app downloaders and nondownloaders (Table 2).

Table 2. Baseline characteristics across levels of engagement.

Baseline characteristics	Consented (N=201)	App nondownloaders (n=101)	App downloaders (n=100)	Completed <level 5 (n=42)	Completed ≥level 5 (n=58)
Age (years), mean (SD)	57.30 (10.92)	58.52 (10.75)	56.06 (11.00)	57.33 (11.58)	55.14 (10.57)
Nonwhite, n (%)	91 (45.3)	53 (52.5)	38 (38.0)	15 (35.7)	23 (39.7)
Female, n (%)	100 (49.8)	51 (50.5)	49 (49.0)	20 (47.6)	29 (50.0)
Weight (kg), mean (SD)	104.5 (26.49)	99.40 (27.58)	106.63 (25.84)	103.82 (26.08)	108.66 (25.70)
BMI (kg/m^2), mean (SD)	35.8 (7.62)	34.82 (7.89)	36.25 (7.50)	35.11 (7.77)	37.09 (7.26)
Hypertension, n (%)	115 (57.2)	58 (57.4)	57 (57.0)	26 (61.95)	31 (53.4)
Hyperlipidemia, n (%)	22 (10.9)	9 (8.9)	13 (13.0)	6 (14.3)	7 (12.1)
Hemoglobin A1c (%), mean (SD)	9.02 (1.17)	9.04 (1.17)	9.00 (1.17)	9.05 (1.22)	8.96 (1.15)

Baseline survey results for the 143 participants (71.1% of consented participants) who filled out the survey are shown in Table 3. The majority of patients were taking medications, and they rated their nonadherence as minimal, reported a moderate amount of physical activity, rated their diabetes self-efficacy as moderate, and rated their diabetes distress as somewhat elevated. These baseline survey results were also compared across app downloaders and nondownloaders. Participants who did not download the app had significantly higher diabetes self-efficacy than participants who downloaded the app (t_{199}=2.62, *P*=.009). There were no other differences regarding baseline survey results between app downloaders and nondownloaders (Table 3).

Table 3. Survey results across levels of engagement.

Survey question	All survey respondents (N=143)	App nondownloaders (n=43)	App downloaders (n=100)	Completed <level 5 (n=42)	Completed ≥level 5 (n=58)
Currently taking medication, n (%)	140 (97.9)	43 (100)	97 (97)	41 (97.6)	56 (96.6)
Extent of medication nonadherence, mean (SD)	1.68 (0.75)	1.60 (0.58)	1.17 (0.82)	1.63 (0.89)	1.78 (0.76)
Self-report physical activity measures, mean (SD)	3.14 (0.69)	2.85 (1.17)	2.55 (1.12)	2.66 (1.03)	2.48 (1.19)
Diabetes self-efficacy scale, mean (SD)	6.72 (1.76)	7.17 (1.71)	6.53 (1.75)	6.73 (1.93)	6.38 (1.61)
Brief Illness Perceptions Questionnaire, mean (SD)	5.28 (1.30)	5.13 (1.38)	5.40 (1.14)	5.13 (1.16)	5.59 (1.11)
Diabetes Distress Scale, mean (SD)	2.09 (1.20)	2.00 (1.27)	2.13 (1.17)	2.14 (1.21)	2.13 (1.15)
Doctor-recommended home glucose test, n (%)	142 (99.3)	43 (100)	99 (99)	41 (97.6)	58 (100)
Self-care: days per week blood glucose checked, mean (SD)	4.13 (2.62)	3.84 (2.64)	4.26 (2.62)	4.12 (2.79)	4.36 (2.51)

Engagement

Overall, 48 of the 201 participants (23.9%) who consented to participate completed all levels and the follow-up survey, and 100 (49.8%) of those who consented to participate downloaded the app. Of the 100 participants who downloaded the app, 16 (16.0%) completed 0 levels, 26 (26.0%) completed 1 to 4 levels, 10 (10.0%) completed 5 to 11 levels, and 48 (48.0%) completed all 12 levels of the app and the follow-up survey.

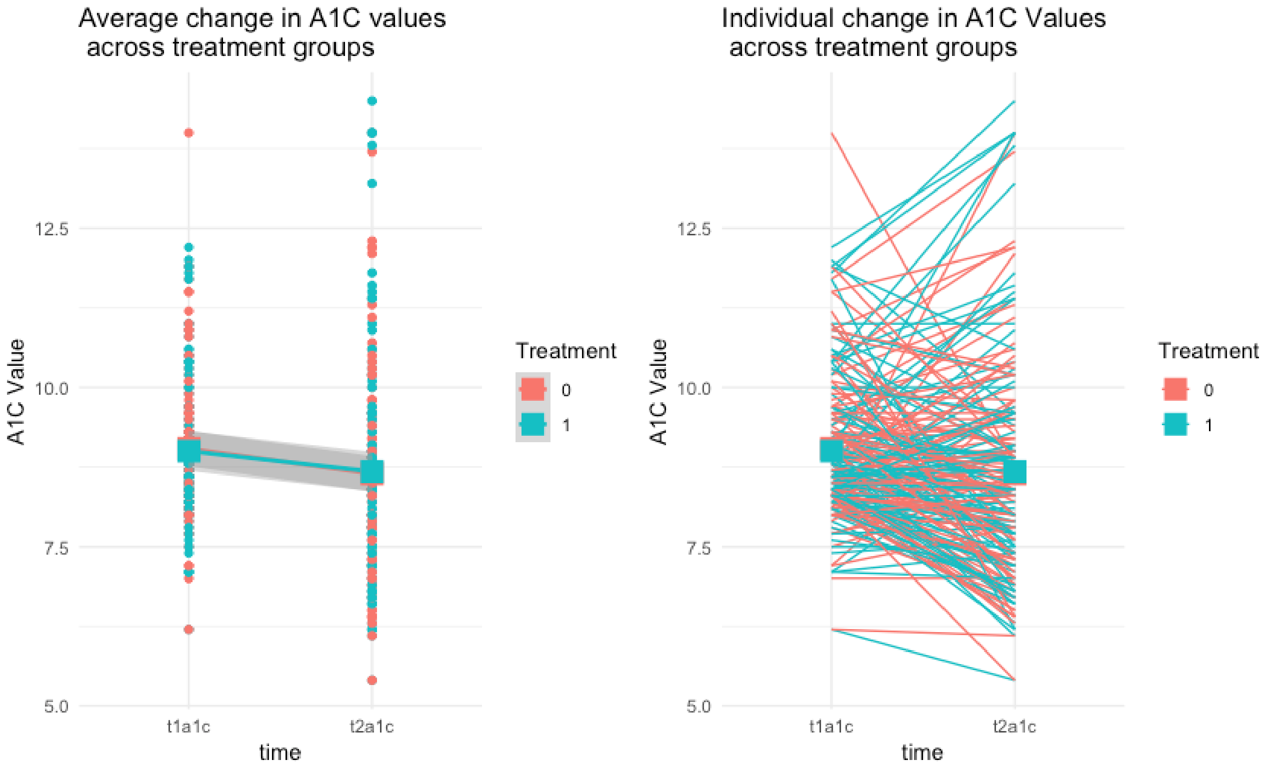
There were no significant or descriptively meaningful differences among participants who completed 0 to 4 levels of the app. Similarly, there were no significant or descriptively meaningful differences among participants who completed 5 to 12 levels of the app (see Table 3). Because of the a priori distinction that the completion of level 5 or above indicates a markedly increased level of engagement, all subsequent analyses related to engagement examined engagement as a dichotomous variable. As shown in Table 2, baseline characteristics were not significantly different across high engagers (completed level 5 or higher) and low engagers (completed 0 to 4 levels). At baseline, high engagers (as compared to low engagers) had more severe perceptions of illness on their Brief Illness Perceptions Questionnaire (t_{92}=2.01, *P*=.047).

Primary Outcome: Change in HbA1c

Figure 1 illustrates the change in HbA1c of all participants, both as an average across app downloaders and nondownloaders (left panel) and as individual trajectories of HbA1c change (right panel).

When adjusting for baseline levels of HbA1c, HbA1c did not change more significantly for the treatment group (–0.32) than for the control group (–0.39; β=.06, *P*=.78). However, higher baseline levels of HbA1c were significantly related to greater decreases in HbA1c (β=–.33, *P*=.001). This significant association is common and is usually assumed to be evidence of regression toward the mean. The null effect of treatment also held when examining a subgroup of participants with a baseline HbA1c greater than 8% (β=.15, *P*=.53). Subgroup analyses of HbA1c and engagement were also conducted. The interaction between baseline HbA1c and the number of levels engaged with in the app (β=.02, *P*=.45) was not associated with change in HbA1c. Thus, when controlling for baseline HbA1c, high versus low engagers did not experience significantly different changes in HbA1c. Additionally, those completing one or more levels of the app (mean pre/post-HbA1c change of –0.41%) did not have significantly different changes in HbA1c compared with those who downloaded the app but completed 0 levels (mean pre/post-HbA1c change of –0.32%; *P*=.73).

Figure 1. Average individual change in hemoglobin A1c (A1C) by group.



Secondary Outcomes: Change in Self-Reported Health Behaviors

Because follow-up measures of self-reported health behaviors were only based on participants who completed the app levels, comparisons in changes in self-reported health behaviors across treatment and control participants could not be performed. However, changes in self-reported health behaviors for the 48 participants completing all levels of the app were examined. Diabetes self-efficacy showed a large and significant increase during app usage (mean change 1.28, *P*<.001, *d*=.83). Severity of illness perceptions showed a small but significant decrease during app usage (mean change –0.51, *P*=.004, *d*=.43). Diabetes distress showed a small but significant decrease during app usage (mean change –0.45, *P*=.006, *d*=0.41). Self-reported medication nonadherence (mean change –0.18, *P*=.11) and physical activity (mean change 0.24, *P*=.14) did not significantly change during app usage. These changes in self-reported health behaviors were consistent across gender (all *P*>.19), race (all *P*>.30), and BMI (all *P*>.27). Increased age was marginally associated with greater improvements in diabetes self-efficacy (β=.03, *P*=.05), and positive hypertension/hyperlipidemia status was associated with greater increases in physical activity (β=.75, *P*=.006).

User Experience/Perceptions of the Program

Those who completed all levels of the app (24% of all consented participants, 48% of app downloaders) rated it highly and would recommend it to others (net promoter score=62.5). Overall, participants who completed the app reported being highly satisfied (mean 4.56, SD 0.76; 1 to 5 scale); that the app has high relevance to their daily lives (mean 4.52, SD 0.68; 1 to 5 scale); that the app helped them better manage their type 2 diabetes (mean 4.38, SD 0.81; 1 to 5 scale); and strongly endorsed that they would recommend this app to colleagues, friends, or family with diabetes (mean 9.06, SD 1.46; 1 to 10 scale). Features of the app that were endorsed as among the most helpful by app completers were the healthy eating articles (n=42, 89%), food comparison tool (n=36, 77%), challenges (n=32, 68%), physical activity articles (n=31, 66%), and blood glucose monitoring articles (n=26, 55%).

Discussion

Principal Results

The Time2Focus mobile app is uniquely designed to assist users in building skills that are needed to reach diabetes self-management goals. Our results demonstrate that participants who used the Time2Focus mobile app showed an improvement in diabetes self-efficacy, which is an essential skill needed in the care and management of diabetes. Users also experienced a significant decrease in severity of illness perceptions. This finding is critically important, since more severe illness perceptions are associated with a stronger cognitive and emotional response to illness, which can negatively influence the perceived ability to cope with the illness [11]. In addition, more severe illness perceptions have also been associated with poorer glycemic control as measured by HbA1c [12].

A decrease in HbA1c in both the intervention and control groups was observed over the 12-week follow-up period. However, there was no difference in change in HbA1c between the intervention and control groups. Importantly, users reported high satisfaction with using the app, found the app to have high relevance to their daily lives, thought the app helped them better manage their diabetes, and said that they would recommend the app to others.

Comparison With Prior Work

Our results mirror many other studies designed to test the effect of technology-based educational apps on HbA1c. For example, a systematic review evaluating mobile apps designed to deliver SMS demonstrated that of 11 studies selected for review, only 45.4% of the studies observed an HbA1c reduction in both the intervention and control groups [5]. Another systematic review that focused on digital health technology and mobile devices for patients with diabetes reviewed studies that were specifically focused on SMS and education for patients with diabetes [13]. Results with respect to change in HbA1c were mixed. The most efficacious studies included personal coaching or personalized messaging, in-person visits, and website components. In contrast, a systematic review performed by Greenwood et al [14] revealed that the majority (18/25) of the 25 review articles and meta-analyses included reported a significant reduction in HbA1c ranging from 0.1% to 0.8%. Communication, education, and feedback are consistently noted to be key design elements for interventions aiming to reduce HbA1c through delivery of diabetes self-management education and support [14,15].

One example of an efficacious multicomponent app is the mySugr mobile app, which was designed to support patients in healthy eating, being active, monitoring and taking medications, risk reduction, problem solving, and healthy coping skills. The features of this app include wireless blood glucose data upload, recording of insulin use and exercise, on-demand direct access to a certified diabetes educator, and algorithms for pattern detection and assistance with day-to-day diabetes management. Results showed that when participants used the mySugr "Bundle" (app plus unlimited test strips and certified diabetes educator–led coaching), there was significant improvement in mean blood glucose (–10.4%), tests in range (+8.5%), and estimated HbA1c (–0.4%) [15].

The lack of difference in HbA1c in this pilot study may be related to our small cohort size as well as our pragmatic approach to HbA1c data collection. Indeed, mobile health technologies that are efficacious when implemented in clinical trials often are not effective when implemented in the field [15,16]. In addition, we chose to utilize clinically derived data because we wanted to test the app and its effects in a real-world environment. It is possible that HbA1c collected by the study team at specific time intervals could have led to a difference in HbA1c in the intervention vs control group.

In addition, differences in socioeconomic status, technological literacy, and low health literacy are known barriers to enrollment and engagement in technology for individuals with diabetes [17,18]. Therefore, it is possible that the lack of difference in HbA1c was secondary to differences in our intervention and control groups with respect to any of the aforementioned characteristics. Lastly, the effects of SMS education can be delayed, and the time frame during which one sees the greatest impact from education and skill-building can be patient-dependent. Indeed, most mobile health trials that have shown effective HbA1c lowering were designed for follow up over an average of 6-12 months [5,14]. Therefore, it is possible that assessing HbA1c after a longer time interval (ie, 6 months) could have led to a different result.

Forty-eight participants (24% of consented participants, 48% of app downloaders) completed all levels of the app, and satisfaction with the app among all users was high. This level of engagement is on par with prior pragmatic trials that tested similar apps, and is particularly high considering that we did not utilize in-person visits or personal coaching components. It is important to note that the number of levels engaged in the app was also not associated with changes in HbA1c (β=–.01, *P*=.76).

Current published data demonstrate low levels of adoption of mobile apps that are designed to assist patients in the management of their diabetes [19]. Further, when individuals do download mobile apps, engagement rates decline over time and attrition rates are often high [19]. Unlike the Time2Focus app, one reason for lack of engagement in other apps is a lack of satisfaction with the design of the app. Our higher level of satisfaction with the app is likely a reflection of the design, which sought to build in real-world situations, allow users to move at their own pace, and is focused on skill-building and improving confidence. The design of the app is consistent with the guiding principles described in the person-based approach to intervention development. These guiding principles describe intervention features that improve acceptability and engagement in an intervention, including promoting user autonomy, competence, and positive experience [6]. Participants also noted they would recommend the app because it was "informative and educational," "keeps up motivation levels," and "holds me accountable."

Strengths

Managing diabetes is time-consuming as well as emotionally and financially burdensome. Having a mobile app option to support diabetes management and education, reduce the severity of illness perceptions, and improve diabetes self-efficacy is a step forward in providing multiple tools to help patients with diabetes achieve their SMS and glycemic goals. The strengths of our pilot study include: (1) the intervention design, which is based on behavior-change theory and focuses on increasing self-efficacy and problem-solving skills; (2) the pragmatic implementation of the intervention and collection of data; and (3) the limited resources needed to implement the intervention.

Limitations

Some limitations of our study are mainly due to the naturalistic setting in which it was conducted, and include a small cohort, the use of naturally occurring HbA1c as the primary outcome, the short duration of the study, and lack of randomization. Other limitations of our study may be addressed in future studies even within naturalistic settings. These limitations include the inherent tying of incentives to intervention completion, and the completion of follow-up surveys in only the 48 participants (23.8%) who completed all levels of the app.

With regard to only administering follow-up surveys to participants who completed the app, it should be noted that the majority of our inferences regarding changes in self-efficacy and other self-reported measures can only be applied to those participants and not to those that dropped out early. Thus, this pilot study largely presents evidence on the amount of engagement across all participants, but changes only within participants willing to engage throughout the app. Although resource constraints influenced our ability to collect additional follow-up data from participants who dropped out, there would be tremendous utility in capturing additional feedback on why some participants dropped out. In this regard, qualitative interview methods may have been instrumental in capturing rich and valuable data on why some individuals disengaged without the need to fully survey all participants who disengaged, which may be prohibitive.

Conclusions

Participants who used the Time2Focus mobile app showed an improvement in diabetes self-efficacy and a decrease in severity of illness perceptions. The decrease in HbA1c observed in app users relative to nonusers during this limited pilot study was not statistically significant. However, uptake and application of lessons learned from SMS may be delayed. Future research is needed to address how to increase engagement in SMS and to investigate if follow up over a longer period would result in a significant change in clinical outcomes such HbA1c.

Acknowledgments

This work was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (BCB, grant number KL2TR001115). DB was supported by Grant No. TPH 21-000 from the Department of Veterans Affairs Office of Academic Affiliations.

Conflicts of Interest

None declared.

References

1. American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2017. Diabetes Care 2018 May;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]
2. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes Self-management Education and Support in Type 2 Diabetes: A Joint Position Statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Clin Diabetes 2016 May;34(2):70-80 [FREE Full text] [doi: 10.2337/diaclin.34.2.70] [Medline: 27092016]
3. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2018 Dec;41(12):2669-2701 [FREE Full text] [doi: 10.2337/dci18-0033] [Medline: 30291106]
4. American Diabetes Association. 5. Lifestyle Management: Standards of Medical Care in Diabetes. Diabetes Care 2019 Jan;42(Suppl 1):S46-S60. [doi: 10.2337/dc19-S005] [Medline: 30559231]
5. Adu MD, Malabu UH, Callander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the Development of Mobile Phone Apps to Support Diabetes Self-Management: Systematic Review. JMIR Mhealth Uhealth 2018 Jul 21;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]
6. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
7. Voils CI, Maciejewski ML, Hoyle RH, Reeve BB, Gallagher P, Bryson CL, et al. Initial validation of a self-report measure of the extent of and reasons for medication nonadherence. Med Care 2012 Dec;50(12):1013-1019 [FREE Full text] [doi: 10.1097/MLR.0b013e318269e121] [Medline: 22922431]
8. Ritter PL, Lorig K, Laurent DD. Characteristics of the Spanish- and English-Language Self-Efficacy to Manage Diabetes Scales. Diabetes Educ 2016 Apr;42(2):167-177. [doi: 10.1177/0145721716628648] [Medline: 26846336]
9. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res 2006 Jul;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]
10. Fisher L, Glasgow RE, Mullan JT, Skaff MM, Polonsky WH. Development of a brief diabetes distress screening instrument. Ann Fam Med 2008;6(3):246-252 [FREE Full text] [doi: 10.1370/afm.842] [Medline: 18474888]
11. Broadbent E, Wilkes C, Koschwanez H, Weinman J, Norton S, Petrie KJ. A systematic review and meta-analysis of the Brief Illness Perception Questionnaire. Psychol Health 2015;30(11):1361-1385. [doi: 10.1080/08870446.2015.1070851] [Medline: 26181764]
12. Mc Sharry J, Moss-Morris R, Kendrick T. Illness perceptions and glycaemic control in diabetes: a systematic review with meta-analysis. Diabet Med 2011 Dec;28(11):1300-1310. [doi: 10.1111/j.1464-5491.2011.03298.x] [Medline: 21418098]
13. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]

14. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
15. Debong F, Mayer H, Kober J. Real-World Assessments of mySugr Mobile Health App. Diabetes Technol Ther 2019 Jun;21(S2):S235-S240. [doi: 10.1089/dia.2019.0019] [Medline: 31169427]
16. van Onzenoort HAW, Menger FE, Neef C, Verberk WJ, Kroon AA, de Leeuw PW, et al. Participation in a clinical trial enhances adherence and persistence to treatment: a retrospective cohort study. Hypertension 2011 Oct;58(4):573-578. [doi: 10.1161/HYPERTENSIONAHA.111.171074] [Medline: 21825228]
17. Wayne N, Ritvo P. Smartphone-enabled health coach intervention for people with diabetes from a modest socioeconomic strata community: single-arm longitudinal feasibility study. J Med Internet Res 2014 Jul 06;16(6):e149 [FREE Full text] [doi: 10.2196/jmir.3180] [Medline: 24907918]
18. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mHealth App Intervention for Persons With Type 2 Diabetes and its Associations With Initial Self-Management: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]
19. Adu M, Malabu U, Malau-Aduli A, Malau-Aduli B. Users' preferences and design recommendations to promote engagements with mobile apps for diabetes self-management: Multi-national perspectives. PLoS One 2018;13(12):e0208942 [FREE Full text] [doi: 10.1371/journal.pone.0208942] [Medline: 30532235]

Abbreviations

EHR: electronic health record
HbA1c: hemoglobin A1c
SMS: self-management support

Edited by G Eysenbach; submitted 18.12.19; peer-reviewed by M Twomey, E Omaki; comments to author 03.04.20; revised version received 28.05.20; accepted 11.11.20; published 20.01.21.

Please cite as:
Batch BC, Spratt SE, Blalock DV, Benditz C, Weiss A, Dolor RJ, Cho AH
General Behavioral Engagement and Changes in Clinical and Cognitive Outcomes of Patients with Type 2 Diabetes Using the Time2Focus Mobile App for Diabetes Education: Pilot Evaluation
J Med Internet Res 2021;23(1):e17537
URL: http://www.jmir.org/2021/1/e17537/
doi:10.2196/17537
PMID:33470947

Original Paper

A Mobile App to Support Clinical Diagnosis of Upper Respiratory Problems (eHealthResp): Co-Design Approach

João Moura[1*], MD; Ana Margarida Pisco Almeida[2*], PhD; Fátima Roque[3,4*], PhD; Adolfo Figueiras[5,6,7*], PhD; Maria Teresa Herdeiro[1], PhD

[1]Department of Medical Sciences, iBiMED—Institute of Biomedicine, University of Aveiro, Aveiro, Portugal

[2]Department of Communication and Art/DigiMedia, University of Aveiro, Aveiro, Portugal

[3]Research Unit for Inland Development, Polytechnic Institute of Guarda (UDI/IPG), Guarda, Portugal

[4]Health Sciences Research Centre, University of Beira Interior (CICS-UBI), Covilhã, Portugal

[5]Department of Preventive Medicine and Public Health, University of Santiago de Compostela, Santiago de Compostela, Spain

[6]Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiology and Public Health-CIBERESP), Madrid, Spain

[7]Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain

[*]these authors contributed equally

Corresponding Author:
Ana Margarida Pisco Almeida, PhD
Department of Communication and Art/DigiMedia
University of Aveiro
campus universitário de santiago
Aveiro, 3810-193
Portugal
Phone: 351 234 370389 ext 2370
Email: marga@ua.pt

Abstract

Background: The misuse of antibiotics is a global public health issue that fosters bacterial resistance and jeopardizes generational health. The development of validated tools such as web-based courses and mobile apps to enhance clinical decisions in upper respiratory infections is of great importance in reducing the incorrect use of antibiotics in these situations.

Objective: The aim of this study was to design and prevalidate the interface of a mobile app to assist and provide clinical support in the diagnosis of upper respiratory problems. We aimed to assess the adequacy and usability of the interface of the tool in the belief that it could be beneficial to health care delivery in the clinical decision setting.

Methods: Using a co-design approach that brought together professionals in interface design and experts in pharmacology and pharmacoepidemiology, the mobile app interface was evaluated through peer review sessions held by interface design professionals on a heuristic survey. The reviewers accessed a high-fidelity interactive mock-up of the interface and filled in a questionnaire to assess the dimensions of *layout and visual design* and *navigation and tasks*. The resulting feedback of this evaluation supported the redesign of the primary interface, which was assessed for the second time by 2 of the previously mentioned reviewers.

Results: With 4 as the highest score, the interface scored a mean of 3.16 (SD 0.45; median of the means 3.2) for *layout and visual design* and a mean of 3.43 (SD 0.33; median of the means 3.51) for *navigation and tasks*, reflecting an overall positive evaluation. The open-ended commentaries allowed us to better understand specific recommendations of the reviewers. Throughout this section, approximately 0.98 comments per parameter were registered, reflecting a high level of effectiveness of the chosen parameters in identifying potential problems. The resultant beta version of the interface, addressing the majority of the detected problems, was further assessed by 2 of the previous reviewers, validating the new design. Future tests with physicians and pharmacists will help assess credibility and user experience dimensions.

Conclusions: Our study revealed that the designed interface is easy to interpret and use. Peer reviewers raised important issues that could be easily fixed and positively reassessed. As a result, the study enabled us to produce a new tool for interface usability assessment and a set of recommendations for developing mobile interfaces for clinical decision support systems in the scope of upper respiratory problems.

(J Med Internet Res 2021;23(1):e19194) doi:10.2196/19194

KEYWORDS

mHealth; Clinical Decision Support Systems; respiratory system; diagnose; interface; mobile phone

Introduction

Background

Antibiotic resistance is a major public health problem worldwide that is mostly fostered by inappropriate use of antibiotic medications. At the same time, data and advances in health care are growing not only in quantity but also in complexity; thus, health systems, practitioners, and even patients are required to be in a constant learning state to achieve effective monitoring and evaluation [1].

With the near total ubiquity of mobile technologies, mobile health (mHealth) is becoming an increasingly established field with important results in different domains [2-4]. This gives rise to new possibilities in enhancing clinical decisions in all medical fields, including antibiotic prescription and dispensing.

In this framework, the project eHealthResp proposes to create and evaluate eHealth tools to support clinical decisions and patient empowerment in the management of upper respiratory infections. These tools include an web-based course targeted toward physicians and pharmacists and a clinical decision support system (CDSS) mobile app targeted toward physicians, pharmacists, and patients.

At a later stage of the research, both tools will be used by participants of an educational intervention about antibiotic prescription and dispensing for upper respiratory symptoms supported by the eHealthResp project.

Expected outcomes include a decrease and improvement in the use of antibiotics as well as a comprehensive list of guidelines in designing and implementing feasible and usable tools for CDSS in a broad scope, particularly for cases involving the upper respiratory system.

In the realm of these predictions, the work in progress and the main issue of this paper focus on the methodology and subsequent results within the interface design stage of the mobile app tool for smartphones running Android or iOS operative systems.

This research stage can be summarized within the following 4 components:

- Goals:
  - The goals of the research are to develop the interface of a CDSS tool that supports decisions while prescribing and dispensing antibiotics in cases involving upper respiratory system symptoms, evaluate its usability, and measure how well the human-computer interaction (HCI) experts in the peer review sessions perceive the interface in terms of effective, efficient, and satisfactory use [5].
- Conceptual framework
  - The incorrect use of antibiotics represents the main worldwide factor for the increasing bacterial resistance to these drugs, requiring a more efficient approach to prescription and dispensing processes.
  - There is a need to use an interdisciplinary approach and co-design methodologies when developing mobile health interfaces.
- Research question
  - How can we develop an interface for an app to support physicians, pharmacists, and patients to properly use previously validated algorithms for upper respiratory symptoms?
- Methods
  - The development process was organized in 4 steps (Figure 1): (1) a primary step regarding a literature search and state-of-the-art analysis, followed by (2) the design of the interface by a team of experts, namely 2 HCI practitioners and 3 experts in pharmacology and pharmacoepidemiology, resulting in the alpha version of the interface. This version went through a prevalidation of the developed interface throughout (3) peer review sessions in a heuristic style evaluation [6] executed by 5 HCI practitioners with experience in interface design and availability. Following the tests, (4) the data analysis stage took into account the experts' outcome and produced the interface redesign (beta version), including a new validation phase, leading to the final proposal.

Figure 1. Design of the research.

eHealthResp Mobile App

Currently targeted toward Android and iOS mobile devices, the eHealthResp mobile app is under development and consists of a diagnostic tool designed to enhance the clinical decision set facing upper respiratory symptoms. The algorithm behind the app and a version of the app were previously validated and designed for the pharmaceutical scope by a research group from the University of Santiago de Compostela based on the work of Molero et al [7]. This served as the basis for the development of the new proposal.

At the end-user level, the tool consists of a wizard that guides the user through the dichotomous key-like algorithm and, in cases in which a disease is identified, presents the end results in a diagnosis format, including *know more*, *treatment*, *prognosis* and *when to derive* information. The workflow was kept linear, straightforward, and with little to no deviations from the main course to keep the user's focus on the diagnosis.

Although the previously developed app, available on the Google Play store, presents the wizard at a functional level, the interface design challenge was not considered when it was developed. At the design level, the major weaknesses of the app concern the use of stretched pictograms, misuse of Android navigation elements (eg, tabs, buttons), lack of consistency between pages and graphical elements (eg, buttons, pictograms, backgrounds), unoptimized image compression, and an overall assessment reflecting an unpleasant interface that fails to entice the user to trust and reuse the tool [8,9].

Methods

Overview and State of the Art

Using the Scopus database, a combination of the keywords *mobile app* (or *mobile health* or *app*), *clinical decision*, and *respiratory* was used to search for similar studies.

Due to the lack of direct references concerning diagnosis apps and design for mHealth, complementary searches were held combining the terms *algorithm*, *mobile*, *design*, *diagnosis*, or *diagnose*, and *respiratory*. From this search, 9 other articles were selected based on title and keywords. In total, 47 articles were selected and analyzed.

In addition to this review, a nonexhaustive benchmark-like search was conducted to gather a glimpse of the state of art in interface and design options for apps in the respiratory and CDSS fields. For this, separate searches were held on the tags *respiratory* and *clinical decision* in the 2 most relevant app stores (Google Play and App Store). For both the tags and stores, a subsearch for paid apps was performed.

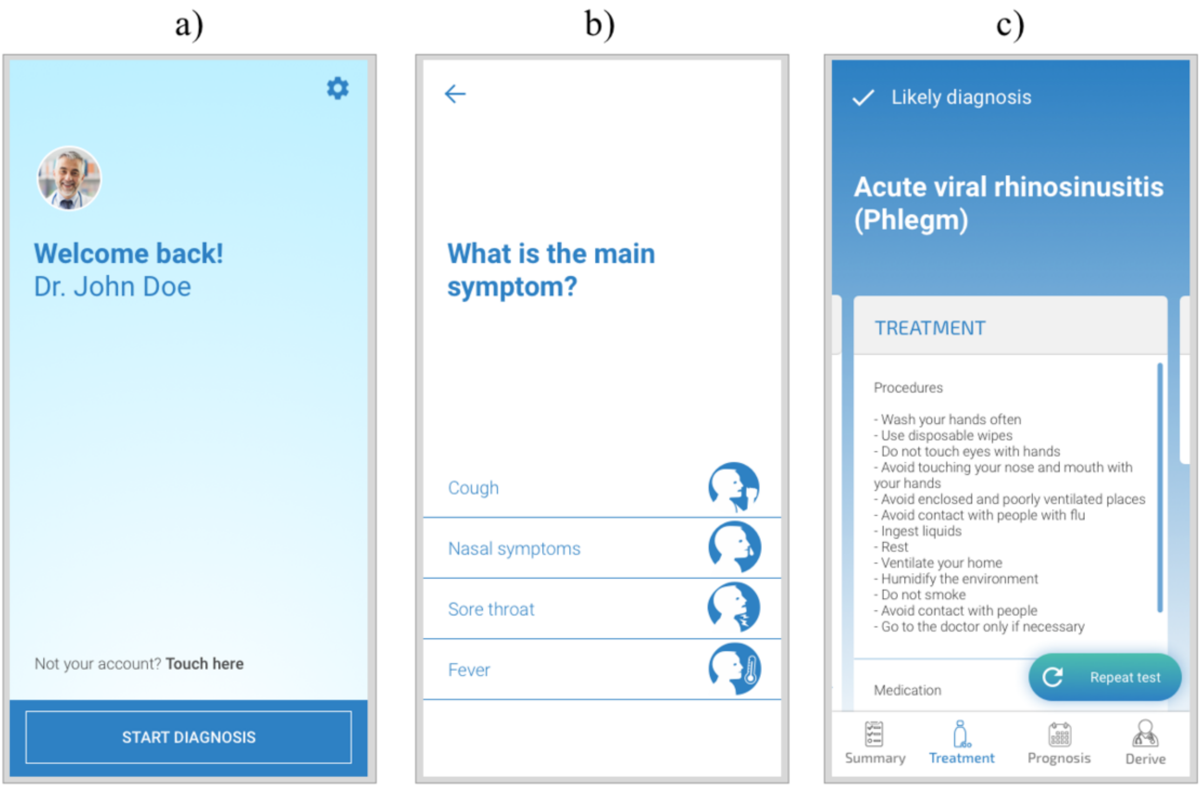
The inclusion criteria were as follows: (1) respiratory system representations, (2) clinical decision tools, (3) quiz formats, and (4) informational or educational content involving the respiratory system. Apps were listed if they included at least one of the previous criteria. Apps that had cross-functionalities with the app under development were highlighted.

For each app, the list included inherent details (title, icon, link, creator, classification, price, description, last update, size, number of downloads, and screenshots) and 2 evaluation lists (pro- and counter-considerations) of the interface characteristics concerning the given screenshots and listed functionalities and based on the literature review. When available for free, highlighted apps were downloaded for a deeper analysis.

Interface Design

To achieve an end design that suits the app users (physicians and pharmacists), the interdisciplinary team involved in this project worked together toward the development of the alpha version of the interface (Figure 2). In addition to periodic meetings, a questionnaire was used to assess the opinions of experts in pharmacology and pharmacoepidemiology about different interface solutions, assigning experience-based values throughout a participatory strategy [10]. This includes understanding the needs and preferences of experts and their peers (physicians and pharmacists) regarding content (script and imagery) and visualization formats. Using fast prototyping tools (Sketch and Marvel App), high-fidelity interface mock-ups were made according to participants' previous feedback and presented to the participants.

Figure 2. Light theme interface design examples before validation. (a) Homepage; (b) first question page; (c) likely diagnosis page.



- Wash your hands often
- Use disposable wipes
- Do not touch eyes with hands
- Avoid touching your nose and mouth with your hands
- Avoid enclosed and poorly ventilated places
- Avoid contact with people with flu
- Ingest liquids
- Rest
- Ventilate your home
- Humidify the environment
- Do not smoke
- Avoid contact with people
- Go to the doctor only if necessary

To limit the time spent in developing high-fidelity mock-ups, the design covered strictly necessary paths to walkthrough the most decisive type-pages. Redundant, secondary, and broadly studied representations were avoided, such as pages with the same structure but different content and the Settings page.

Peer Review Sessions

To attain an outsider point-of-view and opinion, a first series of peer review sessions was held by a group of 5 experts in interface design. This group was conveniently gathered based on the availability and diversity of the members' academic and professional backgrounds.

After an introduction to the research framework and approach, each expert reviewer received a *PDF* document with the interface mock-up pages and an interactive high-fidelity mock-up version of the interface in the web-based app Marvel App. Both the *PDF* document and the Marvel mock-up presented 2 walkthrough paths: one for a light version of the interface, that is, brighter, and the other in a darker mode. To assess the quality of the navigation and graphical design of the interface, a ready-to-fill spreadsheet was provided. This document, an adapted version of the web-assessment tools, *38 page layout and visual design usability guidelines* and *29 navigation and IA usability guidelines* (IA stands for *information architecture*), both part of the "247 web usability guidelines" by David Travis [11]. The guidelines provided by Travis for web usability consist of bullet point lists that are meant to be used as assessment tools to improve consistency and design practices. As the author emphasizes, these guidelines could and should be adapted to the specific context in scope. The assessment tool was translated from English to Portuguese and adapted to keep the focus on the mobile app scope and the 2 main dimensions concerning interface and usability: (1) layout and visual design and (2) navigation and tasks. The adapted version presents a set of 30 parameters to assess layout and visual design and 21 parameters for navigation and tasks. Each parameter presented a statement suggesting a good assessment for a specific criterion (eg, "1. Content density is appropriate for target users and relevant tasks"). The reviewer had the possibility to classify each parameter through an adapted version of the Likert scale, measuring from 1 (Do not agree) to 4 (Totally agree), so that the design team could perceive the importance of the parameters in a more hierarchical way and determine whether to ignore or pay attention to each parameter. The choice for the lack of a fifth neutral middle value was taken to force the reviewer's positioning, as applied in the original binary (complies or does not comply) guideline tool by Travis. Similar to the original tool, each parameter had an open-ended space for observations.

Data Analysis

The mean and median were calculated to interpret the quantitative data collected throughout the reviewers' evaluation using the adapted Likert scale for each bullet point. For the qualitative information gathered through the commentaries, a mean median was calculated to express the number of commentaries by bullet point. Given the relatively small sample of reviewers, every commentary was taken into consideration regarding the specific layout elements that were being referred to and the importance of the issue being addressed.

Results

Overview and State of the Art

The search was conducted between February and April 2019, and a total of 297 references were found. Only 34 articles directly addressed or focused on the searched issues. These 34 articles comprised a publishing period between 2004 and 2019. In total, 9 of the articles found describe studies conducted by the same organization—Allergic Rhinitis and its Impact on Asthma (ARIA) [12-20].

Some of the studies directly addressed the research framework (mobile interface design for CDSS in respiratory conditions); the majority of these focused on the effectiveness of a particular app in enhancing diagnosis and prescription, but few of them consider the user experience with the interface and the hedonic qualities of the interface that influence its long-term use [21]. Nonetheless, they gave important hints for the production of valuable and understandable tools.

The following list presents the main results of the overview, with the articles found and 13 other references that were derived from the original 34:

1. Evidence of mobile phones as potentially useful tools for health purposes [12,14,17,20,22-42].
2. List of state-of-the-art apps in the health field [43-45] and app ratings services [28], further used in the state-of-the-art review.
3. Guidelines for implementing CDSS using smartphones [13,14,17-19,27,30,36,44,46-52].
4. Guidelines for Android design [53] and HCI design [11,46,48,54-59].

The state-of-the-art search covered a period between March 20 and 30, 2019, and produced a list of 98 apps (34 on the App store and 64 on the Play Store. In total, 15 of these apps, were highlighted and 1 was found in both stores).

Overall, the analyzed apps exhibited the following identified characteristics:

- A chat allegory or personal assistant.
- Credibility through minimalism.
- Good visibility of the system.
- Use of validated standards and algorithms.
- Recognizable imagery.
- Use of the system norms.
- Paid apps did not reflect better design or more validated resources.

Interface Design

As already mentioned, this stage produced a new interface proposal featuring one static and one interactive mock-up. This design took into account the guidelines found during the literature search and state-of-the-art survey stage, the CDSS tool requirements expressed through the previously developed app (eRes), and the researchers' considerations. Among the considerations, the practitioners suggested the design of 2 themed versions (*light* and *dark*), for the purposes of personalization and comparison of the visibility of the layout structure. Each version consists mostly of a color switch between the content page and the background color. The secondary or action color was maintained, and the problem or error color was adjusted to improve its visibility.

Further considering the color issue, different chromatic choices were used to visually distinguish the interfaces for different users (physicians, pharmacists, and patients).

To convey the project aesthetic and credibility standards, the original interface layout was broadly redesigned, including reconsideration of the interface elements and their positioning. This includes not only the color aspect but also the typeface, buttons, animations, and redesign of the illustrative pictograms. Small tweaks to the navigation structure were performed on the Results page to gather related information, highlight important content, and hide secondary information in secondary pages.

Peer Review Sessions

The input given was used to produce the beta version and included the reconsideration of gradients (to be avoided), layout (repositioning elements such as questions and answers and buttons such as *Repeat test*), pictograms (redesigned and avoided in some cases), typefaces (avoiding bold and light versions), and the design of a confirmation page before finishing the test.

After the redesign, the new interface was presented in person through meetings with 2 of the reviewers, who were selected based on the amount of input given previously. These meetings were also intended to ensure that the design team correctly interpreted their responses and assessed whether the new proposals effectively rectified the issues that were found.

Data Analysis

With 4 as the maximum value of the abovementioned Likert scale, the interface scored a mean of 3.16 (SD 0.45; median of the means 3.2) for layout and visual design and a mean of 3.43 (SD 0.33; median of the means 3.51) for navigation and tasks (the data are presented in detail in Tables 1 and 2). Within the chosen criteria, these results can be acknowledged as confident positive evaluations of the proposed interface.

Table 1. Layout and visual design checklist and respective peer review test results.

Checklist item		Evaluation						
		T[a]1	T2	T3	T4	T5	Mean (SD)	Median
1	Displayed content density is appropriate for target users and their tasks	3	4	2	4	4	3.4 (0.89)	4
2	The layout helps to keep the focus of attention on what to do next	3	3	2	2	4	2.8 (0.84)	3
3	There is a clear “starting point” for each screen	3	3	3	3	4	3.2 (0.45)	3
4	The app is pleasant to look at	3	3	2	3	3	2.8 (0.45)	3
5	The app has a consistent and clearly recognizable appearance that will interest users	4	3	2	3	4	3.2 (0.84)	3
6	The different app screens share a consistent layout	3	4	3	3	4	3.4 (0.55)	3
7	Related information and functions are grouped and clearly recognizable	4	4	2	3	4	3.4 (0.89)	4
8	The screens respect a grid of horizontal and vertical alignments	4	4	3	3	4	3.6 (0.55)	4
9	There is a good balance between information density and white space	4	3	3	2	3	3.0 (0.71)	3
10	Colors work well together, and the use of complicated backgrounds is avoided	3	4	2	3	3	3.0 (0.71)	3
11	Colors are used to structure and group items	4	2	2	3	2	2.6 (0.89)	2
12	The use of contrasting elements (eg, bold text) is applied to emphasize important topics/ or categories	4	4	2	2	3	3.0 (1)	3
13	The screens are organized well and have no irrelevant information	4	4	4	3	4	3.8 (0.45)	4
14	Icons, pictograms, and graphics are recognizable and/or intuitive to understand (concrete and familiar)	3	3	2	4	4	3.2 (0.84)	3
15	The basic elements (screen titles, navigation items...) are easy to find	4	4	4	4	2	3.6 (0.89)	4
16	Attention-grabbing elements (eg, animations, bold colors, distinctive sizes) are used with caution and only when needed	2	3	3	2	2	2.4 (0.55)	2
17	Icons are visually and conceptually distinct but share a common harmony (clearly part of the same family)	3	4	2	1	4	2.8 (1.3)	3
18	Clickable contents (buttons) are clearly recognizable as such	3	3	3	3	4	3.2 (0.45)	3
19	The relationship between controls and their actions is obvious	3	4	4	3	4	3.6 (0.55)	4
20	Radio buttons and check boxes are used appropriately	1	4	3	3	3	2.8 (1.1)	3
21	Nonbutton items do not have button characteristics	2	4	4	3	3	3.2 (0.84)	3
22	Clickable items and content (buttons) include redundant labels or subtitles	1	3	1	1	3	1.8 (1.1)	1
23	The most important information is clearly displayed in the start zone (no need to scroll)	1	4	4	3	4	3.2 (1.3)	4
24	The app clearly shows when there is off-screen content that requires scrolling to view	1	2	4	3	3	2.6 (1.14)	3
25	Meaningful labels, functional background colors, and the use of margins and white space help the user identify distinct items	4	3	3	4	4	3.6 (0.55)	4
26	Typeface fonts are used consistently	4	3	3	3	4	3.4 (0.55)	3
27	Text fonts (typeface) are readable	2	4	3	4	3	3.2 (0.84)	3
28	Use of italic text is avoided	4	4	3	4	4	3.8 (0.45)	4
29	The app avoids extensive use of capitalized text	4	4	3	4	4	3.8 (0.45)	4
30	Textual content is neither too short nor too long	4	3	3	4	3	3.4 (0.55)	3

[a] T: Test.

Table 2. Navigation and tasks checklist and respective peer review test results.

Checklist item		Evaluation						
		T[a]1	T2	T3	T4	T5	Mean (SD)	Median
1	The app has no irrelevant, unnecessary, or distracting information	4	4	4	4	3	3.8 (0.45)	4
2	Excessive text, animations, or images have been avoided	4	3	4	4	4	3.8 (0.45)	4
3	The user does not need to use memory to scroll through the app	4	3	4	4	4	3.8 (0.45)	4
4	The main path is clear, avoiding distractions	4	3	4	4	3	3.6 (0.55)	4
5	The information is presented in a simple and natural way	4	4	2	3	3	3.2 (0.84)	3
6	The number of screens per task has been minimized	4	4	3	3	4	3.6 (0.55)	4
7	The app requires minimal scrolling and clicks	3	3	4	3	3	3.2 (0.45)	3
8	The app correctly anticipates the user's next intentions	3	3	4	3	2	3.0 (0.71)	3
9	The use of metaphors is understandable	3	3	4	3	3	3.2 (0.45)	3
10	If there is an image or icon alongside a button, it is relevant for the task	3	4	3	4	4	3.6 (0.55)	4
11	Commands and actions are presented as buttons or gestures, not as hyperlinks	4	3	4	3	4	3.6 (0.55)	4
12	A new user can use the most common functions without assistance	3	4	4	3	4	3.6 (0.55)	4
13	There is a convenient and obvious way to go through the different screens of the app	4	4	3	3	4	3.6 (0.55)	4
14	The most relevant information is easily accessible	2	4	3	3	4	3.2 (0.84)	3
15	Navigation is organized in the most logical way and oriented to the app’s tasks	4	4	4	3	2	3.4 (0.89)	4
16	The structure of the app is simple and without unnecessary levels	4	4	4	3	4	3.8 (0.45)	4
17	The main sections of the app are available from any screen and there are no dead ends	3	2	4	2	2	2.6 (0.89)	2
18	Navigation feedback is appropriate	3	3	4	3	3	3.2 (0.45)	3
19	The app has its own consistent graphic terminology and conventions between the different screens	3	4	2	3	3	3.0 (0.71)	3
20	Only navigation screens (such as the homepage) can be viewed without scrolling	3	4	2	4	4	3.4 (0.89)	4
21	The app allows the user to browse at their own pace	3	4	4	4	4	3.8 (0.45)	4

[a] T: Test.

Nonetheless, some parameters scored moderately high values: *22. Clickable items and content (buttons) include redundant labels/subtitles* for layout and visual design was the lowest scored item, with a mean evaluation score of 1.8 (median 1, SD 1.1). Other low-score parameters include *11. Colors are used to structure and group items* (mean 2.6, median 2, SD 0.89) and *16. Attention-grabbing elements (eg, animations, bold colors, different sizes) are used with caution and only when necessary* (mean 2.4, median 2, SD 0.55).

The lowest score for the navigation and tasks scope concerned the parameter *17. The main sections of the app are available from any screen and there are no dead ends* (mean 2.6, median 2, SD 0.89). The following lowest value already achieved an appreciably positive value of 3/3 for *8. The app correctly anticipates the user’s next intentions*.

Although the quantitative approach provides important hints on what to look for and an overall assessment, the open-ended commentaries enable better understanding of the reviewer’s concerns. In this section, layout and visual design received a mean of 1.2 comments per parameter (median 1); the abovementioned parameter 16 was the one that received the most feedback, with comments from 4 of the 5 peer reviewers. All comments for this parameter warned about different issues such as text sizes, contrasts, visual weight, and sparse use of colors. The parameter *11. Colors are used to structure and group items* had 3 comments, all of which highlighted the absence of a more variable color palette (eg, “Color variability is not something that goes into this app”). In addition, with 3 comments, the parameter *20. Radio buttons and check boxes are used appropriately* raised issues concerning the interaction limitations of the mockup.

The navigation and tasks section received approximately 0.714 comments per parameter (median 1). The most frequently commented parameter (3 comments) was *17. The main sections of the app are available from any screen and there are no dead ends*. The comments reflected on the one-way-path aspect of the app, the limitation of the mock-up in turning back one action, the lack of a submission confirmation page (“I would say that before presenting the likely diagnosis there should be an confirmation of intention to ‘submit’”) and the difficulty in locating the button to go back to the homepage and repeat the test (“I can’t find a way to go back to something that resembles the homepage”).

Discussion

Principal Findings

In this study, we present the main stages and outcomes of the current developments in the interface design for the eHealthResp mobile app as a means to produce guidelines for mobile interface development for other CDSS tools with similar characteristics. Guidelines such as these, though available, are still scarce and lack validation [60].

The literature review highlighted the possibility and need for mHealth solutions to enhance diagnosis [13,26-28,31,33,35,40,42,50,52,60]. It gave a glimpse of the large number of mobile apps currently available within the health topic [43-45] and also flagged the lack of cohesive evaluation standards among them [25,26,28,38,41,60]. However, it helped collect important guidelines to foster better CDSS and mHealth solutions [13,15,18,19,27,30,36,44,46-52,61] to positively impact the quality of care regarding diagnosis [47] and potentially support overburdened medical education programs, promoting better patient care [30] and better, quicker, and more confident clinical decision processes by physicians [35].

As one of the most prominent studies found in the literature, Mobile Airways Sentinel networK (MASK), part of the ARIA initiative, focuses on the design and implementation of tools and guidelines for tool development in the scope [13,15,16,18-20,61].

Within several references to the ARIA project, Courbis et al [19] described a cascade-like methodology for implementing clinical decision support from paper guidelines to the MASK mobile app. A similar approach was adopted in building the eHealthResp app, including collaborative ways of designing and evaluating the solution and transforming the validated algorithms into a user-friendly interface.

The eHealthResp mobile interface design also follows a very similar design methodology to that adopted in mPneumonia [46]. The study makes use of prevalidated algorithms transforming them from paper into a step-by-step, user-friendly assessment questionnaire for mobile interfaces. The team also focused on the feasibility and usability, and unlike our study, they managed to gather acceptability levels. Most of the problems raised in the mPneuomonia project were conveniently approached while designing the first version of the eHealthResp mobile interface and were not raised during the appraise by the peer reviewers.

In accordance with the literature review, the state of the art review reiterates the existence of many apps available within the health care spectrum [43-45]. Despite this, these apps are presented through poor classification and evaluation systems [28], making it difficult to search for trustworthy and easy-to-use apps for a specific issue such as CDSS for upper respiratory infections.

Alongside conclusions by Panesar et al [50], we believe that a well-designed and accepted smartphone app can increase awareness of the importance of antimicrobial stewardship and influence some prescribing behaviors. The right information in the right context can reduce uncertainty, particularly in the antibiotic prescription realm [32].

Within the process of designing the interface, the most important factor was the design team's interdisciplinary dimension. We tried to convey the warnings from researchers such as Litvin [49] and Rawson [51] about the need to predict the tool inclusion in the clinical workplace and grant the perception of usefulness to assist in decision making. This was done by directly involving the clinical scope throughout the design study from an early stage. This involvement was granted by an interdisciplinary design team with experienced members in the clinical field and also by targeting the app toward the pharmacists and physicians' participants of the seminar (introduced in the Background section of this paper).

Despite confirming the beneficial outcomes for patients, Terry [28] flags the ongoing issue of classifying and rating mobile apps for health and calls for the inclusion of physicians, patients, and caregivers in the evaluation process. In addition, while detecting strengths, weaknesses, opportunities, and threats of smartphone-supported diagnosis for the particular case of allergy diseases, Pereira [26] highlights the lack of validation for this type of tool (for diagnostic decisions) and calls for multidisciplinary studies, similar to the research in focus on this paper, to obtain high-quality and useful tools.

These conclusions go along with the outcomes expressed throughout the design process, during which insightful outcomes arose from several meetings and questionnaires regarding preferences and worries predicted by the team members closer to the end-user's community. Although some authors call for automated evaluation tools [36], we retained the use of questionnaires and meetings with experts to evaluate the designed app as a pragmatic way of making quick assessments and fostering the interaction between the design and pre-evaluation processes.

In addition, and because the contexts of use and users were already clearly predefined and represented among the research team members, it was possible to address the issues with a close to *contextual design* approach without the need for deep or direct research within the broad and complex scope of the study and all the limitations that this approach could imply [25].

At the same time, fast, interactive, and high-fidelity prototyping was revealed to be a major key factor for the co-design approach. This allowed other nonexpert designers to understand the approach almost seamlessly without requiring any kind of abstraction or written descriptions. Problems arose only with specific limitations of the prototyping tools, such as the inability to conveniently represent horizontal scroll, specific content animations (eg, icons mutations, element dislocations), and different screen aspect ratios.

Other interface design insights were taken into consideration; for example, Martínez-Pérez et al [44] highlight the need to avoid the use of text-only interfaces, making use of the interactivity, images, and logical decision trees throughout algorithms in a step-by-step approach to input data and restrict the input need to the minimum, reducing the time required to complete the diagnosis.

Guidelines for Android interfaces [53] were used to convey consistency with the primary system (Android) in which the mobile app will run. After the app is fully developed, convenient adaptations will be made so that the interface can be used equally in the iOS ecosystem [62].

In the scope of web design, Lindgaard [54] hints at the importance of immediately perceived esthetics, beauty, and visual appeal to grant hedonic values and urge the user to trust and use the tool. However, in the realm of web design, we believe that these values can be adopted in any design project that relies on the visual sense to obtain the user's trust.

In a study by Shneiderman, the "golden rules of interface design" [59] were addressed throughout research by Gong and Tarasewich [55] to re-adapt these rules into "guidelines for handheld mobile device interface design." These guidelines comprise original, adapted, and new guidelines that enable frequent users to use shortcuts; offer informative feedback; design dialog to yield closure; support internal locus of control, consistency, reversal of actions, error prevention and simple error handling; reduce short-term memory load; design for multiple and dynamic contexts; design for small devices; design for limited and split attention; design for speed and recovery; design for top-down interaction; allow for personalization; and design for enjoyment.

Nielsen and Budiu [58] highlight the general characteristics of mobile human-computer interactions. Among other things, they underline the importance of a clear start-up screen and the consistency between app pages and branding.

The authors also bridged the design stage to the evaluation stage, often linking these stages in a circular manner that fosters redesigning and re-evaluation. Authors such as Kushniruk et al [48], Shneiderman et al [59], authors from the Interaction Design Foundation [56], Nayebi et al [57], Nielsen and Budiu [58], and Travis [11] stress the importance of good usability testing, qualitative research, and methodologies, including usability heuristics with well-tested design principles for inspection, walkthroughs, action research methods, and concepts such as affordance.

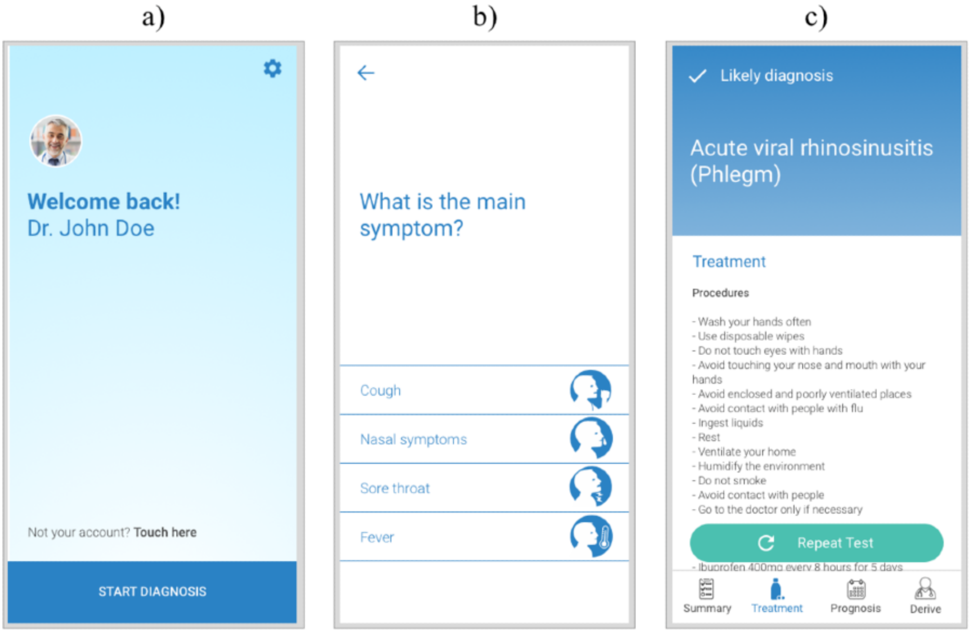
Despite all the evidence that mobile devices are valuable tools for clinical decision-making by both physicians and pharmacists, there is still a need for rigorous evaluation, validation, and best practices for development to ensure the end-quality and safety of the tool [33]. Although the literature was very positive regarding the use of mobile technologies, it also warned about user anxiety issues, limited access to technologies for some, and security concerns [37].

A peer review was necessary to understand the interface limitations in a broad scope. The adopted methods and tools were revealed to be useful and suitable to quickly assess the reviewers' opinions about the interface issues. Overall, the 5 reviewers provided approximately 0.96 comments per parameter, reflecting considerable efficiency of the chosen parameters in raising potential problems. As the main research goal was to detect potential issues rather than to conduct an overall evaluation, the adaptation of the guidelines by Travis [11] enabled a quick and efficient evaluation setup. The shortening of the classic 5-point Likert scale to 4 points allowed a relatively short sample of peer reviewers to express their opinions in a more binary (positive or negative) way for each guideline. The downside of this new scale concerns the negative assessment of some bullet points that could otherwise be classified as neutral evaluation and the overlooking of some positive points. Despite this concern, the reduced number of reviewers and the need to interpret every comment made allowed us to carefully assess each bullet point.

The results supported the creation of the beta version of the interface (Figure 3), addressing most of the detected problems. The version was further reassessed by 2 of the previous reviewers, validating the new design. In this second stage of validation, performed by 2 of the more critical reviewers, the interface was classified as clear and easily usable.

Overall, the processes of co-design, evaluation, redesign, and re-evaluation produced valuable outcomes, addressing major problems of the initial design and proving the processes to be an efficient strategy to speed up the design process. As the main outcome, the study enabled us to create a guidebook for the development, with recommendations (Figure 4) explaining the rationale behind the design choices and the constructive rules for the interface. This guidebook allows developers to have a clear perception of the composition of the interface, summarizing the following chapters: (1) introduction, (2) layout basic elements, (3) page types, (4) specific layout behavior, (5) grid or relative distances, (6) themes, (7) color scheme, (8) typeface, (9) list of pictograms, (10) transitions between pages, and (11) animations.

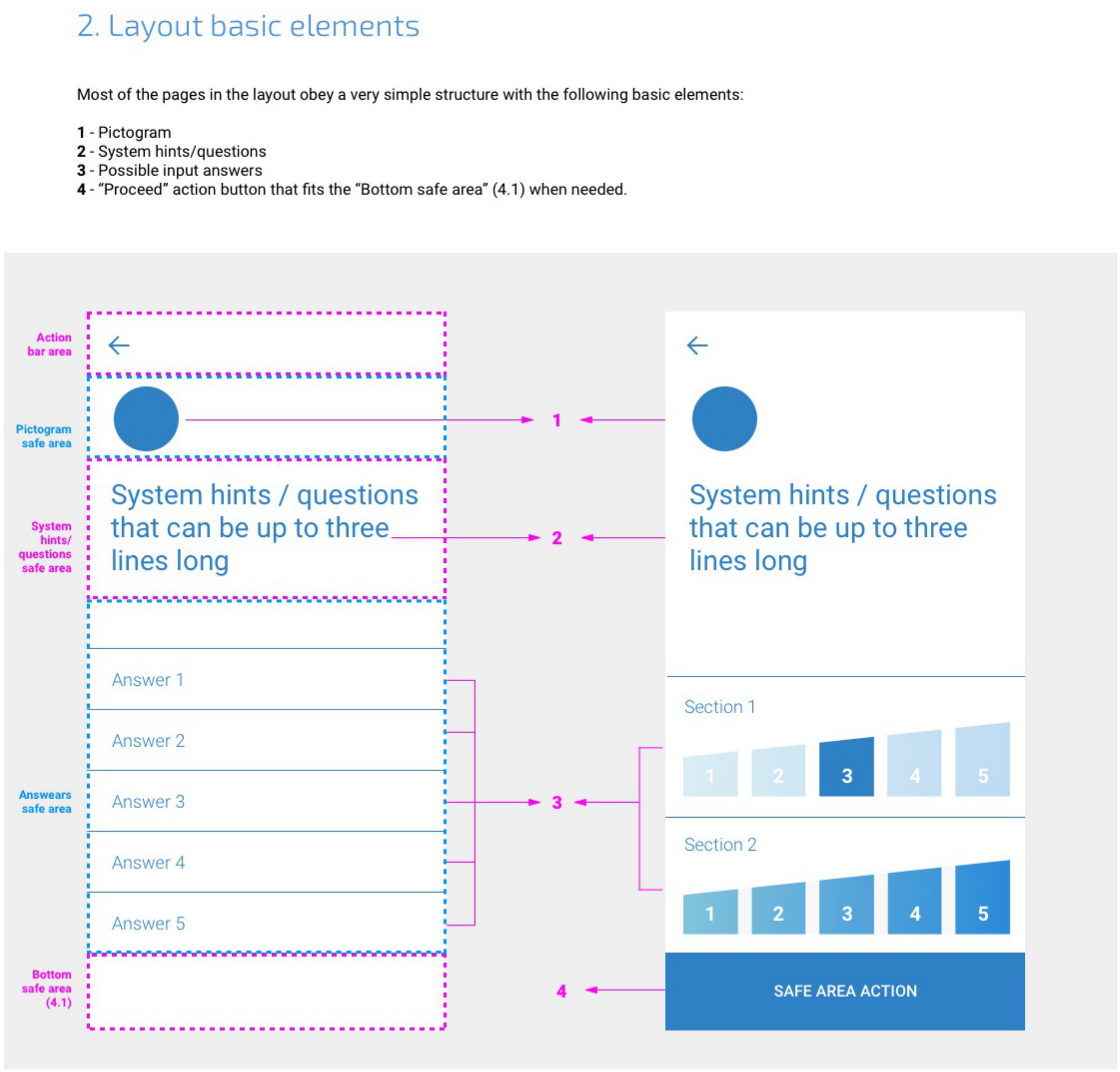
Figure 3. Light theme interface redesign examples after evaluation. (a) Homepage; (b) first question page; and (c) likely diagnosis page. Note: English versions made specifically for the purposes of this paper.



Procedures

- Wash your hands often
- Use disposable wipes
- Do not touch eyes with hands
- Avoid touching your nose and mouth with your hands
- Avoid enclosed and poorly ventilated places
- Avoid contact with people with flu
- Ingest liquids
- Rest
- Ventilate your home
- Humidify the environment
- Do not smoke
- Avoid contact with people
- Go to the doctor only if necessary
- Ibuprofen 400mg every 8 hours for 5 days

Figure 4. Example page of the design-development guidebook (Chapter: Layout basic elements).



2. Layout basic elements

Most of the pages in the layout obey a very simple structure with the following basic elements:

1 - Pictogram
2 - System hints/questions
3 - Possible input answers
4 - “Proceed” action button that fits the “Bottom safe area” (4.1) when needed.

Limitations

The major limitation of the study is the relatively small design team. Although the small team size contributed to speeding up the decision processes, including only 2 HCI experts and 3 experts in pharmacology and pharmacoepidemiology could preclude the representation of the variety of possible contexts of use.

As already highlighted in the previous sections, limitations within the interactive mock-up also presented an issue to the nondesigner team members and reviewers’ assessment in independently interpreting the solution. Specific limitations of the prototyping tools in representing items such as horizontal scrolling, specific content animations (eg, icons mutations, element dislocations), and different screen aspect ratios, caused some confusion in interpreting the solution. In addition, the fact that some buttons being represented were not interactive and led to no alternative path caused some uncertainty regarding the true meaning of these elements among the reviewers and nondesigners.

These limitations could only be clarified in person and within the second evaluation stage with the reviewers.

Future Work

Future work should include the implementation of the app and its confrontation with established assessment criteria such as mHealth evidence reporting and assessment [63].

After this step, usability tests with end-user input must be performed. Conveniently selected physicians and pharmacists in their workplaces will assess the usability of the mobile app using the System Usability Scale as a tool. Their feedback and adherence will better ascertain the effectiveness of all solutions [41].

After the design is revised and established, the research will undertake a pilot study to evaluate the effectiveness of the aforementioned tools, covering 20 primary care physicians, 20 community pharmacists, and 50 patients selected by key informants. This study will gather quantity and quality indicators as response variables within the context of antibiotic consumption to be statistically analyzed on an intention-to-treat basis. The tool should undertake a validation, in which the outcomes from the use of the mobile tool will be compared with those within the usual clinical decision setting (without the tool) together with a user survey regarding the user experience and usefulness of the app [23].

Conclusions

Understanding the true impact of mHealth tools is still an uncertain task, as Forrest et al [39] concluded from their analysis of CDSS solutions. At the time of the study, it was still difficult to perceive the true impact of these tools on patient health outcomes. Although these solutions can significantly improve adherence to antibiotic prescription guidelines, providing easy access to these tools may not be sufficient to achieve higher levels of adherence [34].

As deepened in the literature and state of the art reviews, there is a need to develop recognizable standards in the development of mHealth solutions for CDSS in upper respiratory symptoms control. For this, the main factors influencing the success of these tools must be identified to complement the existing guidelines for mobile development [13,42,44,50-55]. More specifically, there is a need to define requirements relating to layout and content design for usability, acceptability, and usefulness of the app contents and features.

In an attempt to answer the question “How do we develop an interface for an app to support physicians, pharmacists, and patients to properly use previously validated algorithms for upper respiratory symptoms?” the research performed to date has contributed to clarifying some relevant aspects. The use of state-of-the-art tools for high-fidelity prototyping can be crucial to speed up the design process for a multidisciplinary team, not only because it can clearly represent the designers’ conventions to the rest of the team but also because it can help to easily and quickly integrate ongoing suggestions, allowing for a highly interactive co-design process that conveys the team’s concerns.

The developmental methodology enabled us to produce a set of guidelines or templates to produce an app that conveys the requirements of the app and aids the implementation stage.

Acknowledgments

Project PTDC/SAU-SER/31678/2017 was supported by the operational program of competitiveness and internationalization, in its FEDER/FNR component POCI-01-0145-FEDER-031678, the Foundation for Science and Technology, in its state budget component (OE), and the Institute of Biomedicine (iBiMED; UIDB/04501/2020 and POCI-01-0145-FEDER-007628). The main objective of this project is to develop and validate eHealth tools supporting clinical decision-making, focusing on serious public health issues of antibiotic consumption and resistance.

Conflicts of Interest

None declared.

References

1. Best care at lower cost. In: The Path to Continuously Learning Health Care in America (2013). Washington, DC: National Academies Press; 2013.
2. Almeida A, Marques A, Caixinha H. Exercit@rt. In: Encyclopedia of e-Health and telemedicine. Pennsylvania, United States: IGI Global; 2016:1-14.
3. Ribeiro N, Moreira L, Almeida A, Santos-Silva F. Smartphones: innovative tools in cancer prevention. In: Encyclopedia of E-Health and Telemedicine. Pennsylvania, United States: IGI Global; 2016:1041-1058.
4. Ryu S. Book review: mHealth: new horizons for health through mobile technologies: based on the findings of the second global survey on ehealth (global observatory for ehealth series, volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]
5. Ergonomics of human-system interaction — Part 11: usability: definitions and concepts. ISO 9241-11:2018. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2020-12-29]
6. Nielsen J. Usability inspection methods. In: CHI '94: Conference Companion on Human Factors in Computing Systems. 1994 Presented at: ACM Conference on Human Factors in Computer Systems; April, 1994; Boston Massachusetts USA p. 413-414. [doi: 10.1145/259963.260531]
7. Molero GJ, Carandell JM, Cots YJ, Llor C, Muñoz GJ, Paredes SJ. Manual de enfermedades infecciosas en atención primaria. SemFYC. 2010. URL: https://axon.es/ficha/libros/9788496761988/manual-de-enfermedades-infecciosas-en-atencion-primaria [accessed 2020-12-29]
8. Rohrer C. Desirability studies: measuring aesthetic response to visual designs. XDStrategy. 2008. URL: https://www.xdstrategy.com/desirability-studies/ [accessed 2020-12-29]
9. Sousa APD, Almeida AM. Habits and behaviors of e-health users: a study on the influence of the interface in the perception of trust and credibility. Procedia Computer Science 2016;100:602-610. [doi: 10.1016/j.procs.2016.09.201]

10. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
11. Travis D. 247 web usability guidelines. Userfocus Ltd. 2016. URL: http://www.userfocus.co.uk/resources/guidelines.html [accessed 2020-12-29]
12. Bédard A, Basagaña X, Anto JM, Garcia-Aymerich J, Devillier P, Arnavielhe S, MASK study group. Mobile technology offers novel insights into the control and treatment of allergic rhinitis: The MASK study. J Allergy Clin Immunol 2019 Jul;144(1):135-143. [doi: 10.1016/j.jaci.2019.01.053] [Medline: 30951790]
13. Bousquet J, Schunemann HJ, Fonseca J, Samolinski B, Bachert C, Canonica GW, et al. MACVIA-ARIA Sentinel NetworK for allergic rhinitis (MASK-rhinitis): the new generation guideline implementation. Allergy 2015 Nov 13;70(11):1372-1392. [doi: 10.1111/all.12686] [Medline: 26148220]
14. Bousquet J, van Hage M. Correction to: ARIA 2016: Care pathways implementing emerging technologies for predictive medicine in rhinitis and asthma across the life cycle. Clin Transl Allergy 2019 Oct 10;9(1):53 [FREE Full text] [doi: 10.1186/s13601-019-0289-0] [Medline: 31624536]
15. Bousquet J, MASK study group. Correction to: MASK 2017: ARIA digitally-enabled, integrated, person-centred care for rhinitis and asthma multimorbidity using real-world-evidence. Clin Transl Allergy 2019 Oct 9;9(1):51 [FREE Full text] [doi: 10.1186/s13601-019-0290-7] [Medline: 31624534]
16. Bousquet J. Allergic Rhinitis and its Impact on Asthma (ARIA). Clin Exp All Rev 2003 Feb;3(1):43-45. [doi: 10.1046/j.1472-9725.2003.00063.x]
17. Bousquet J, Devillier P, Anto JM, Bewick M, Haahtela T, Arnavielhe S, MACVIA working group. Daily allergic multimorbidity in rhinitis using mobile technology: a novel concept of the MASK study. Allergy 2018 Aug 24;73(8):1622-1631. [doi: 10.1111/all.13448] [Medline: 29569295]
18. Caimmi D, Baiz N, Tanno LK, Demoly P, Arnavielhe S, Murray R, MASK Study Group. Validation of the MASK-rhinitis visual analogue scale on smartphone screens to assess allergic rhinitis control. Clin Exp Allergy 2017 Dec 11;47(12):1526-1533. [doi: 10.1111/cea.13025] [Medline: 28886234]
19. Courbis A, Murray RB, Arnavielhe S, Caimmi D, Bedbrook A, Van Eerd M, et al. Electronic Clinical Decision Support System for allergic rhinitis management: MASK e-CDSS. Clin Exp Allergy 2018 Dec 20;48(12):1640-1653. [doi: 10.1111/cea.13230] [Medline: 29999223]
20. Larenas-Linnemann D, Mullol J, Ivancevich JC, Antó JM, Cardona V, Dedeu T, et al. [MASK (Mobile Airways Sentinel Network). ARIA's comprehensive solution for mobile app for the multimorbidity of allergic rhinitis and asthma]. Rev Alerg Mex 2019 Apr 04;66(1):140-146. [doi: 10.29262/ram.v66i1.578] [Medline: 31013416]
21. Kujala S, Roto V, Väänänen-Vainio-Mattila K, Karapanos E, Sinnelä A. UX Curve: a method for evaluating long-term user experience. Interacting with Computers 2011 Sep;23(5):473-483. [doi: 10.1016/j.intcom.2011.06.005]
22. McGregor JC, Weekes E, Forrest GN, Standiford HC, Perencevich EN, Furuno JP, et al. Impact of a computerized clinical decision support system on reducing inappropriate antimicrobial use: a randomized controlled trial. J of the Amer Med Inf Ass 2006 Jul 01;13(4):378-384. [doi: 10.1197/jamia.m2049]
23. Meyer A, Thompson P, Khanna A, Desai S, Mathews BK, Yousef E, et al. Evaluating a mobile application for improving clinical laboratory test ordering and diagnosis. J Am Med Inform Assoc 2018 Jul 01;25(7):841-847 [FREE Full text] [doi: 10.1093/jamia/ocy026] [Medline: 29688391]
24. Mickan S, Atherton H, Roberts NW, Heneghan C, Tilson JK. Use of handheld computers in clinical practice: a systematic review. BMC Med Inform Decis Mak 2014 Jul 06;14:56 [FREE Full text] [doi: 10.1186/1472-6947-14-56] [Medline: 24998515]
25. Peng Q. Research of a m-health app design for information management of MDTMs. In: Human-Computer Interaction. Interaction in Context. Switzerland: Springer, Cham; 2018:72-82.
26. Pereira AM, Jácome C, Almeida R, Fonseca JA. How the smartphone is changing allergy diagnostics. Curr Allergy Asthma Rep 2018 Oct 25;18(12):69. [doi: 10.1007/s11882-018-0824-4] [Medline: 30361774]
27. Tamposis I, Pouliakis A, Fezoulidis I, Karakitsos P. Mobile platforms supporting health professionals: need, technical requirements, and applications. In: Medical Imaging: Concepts, Methodologies, Tools, and Applications. Pennsylvania, United States: IGI Global; 2016:1020-1043.
28. Terry K. Prescribing mobile apps: what to consider. Med Econ. 2015. URL: https://www.medicaleconomics.com/medical-economics/news/prescribing-mobile-apps-what-consider [accessed 2020-12-29]
29. Jiang Y, Sereika SM, DeVito Dabbs A, Handler SM, Schlenk EA. Using mobile health technology to deliver decision support for self-monitoring after lung transplantation. Int J Med Inform 2016 Oct;94:164-171 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.07.012] [Medline: 27573324]
30. Kabanda S, Rother H. Evaluating a South African mobile application for healthcare professionals to improve diagnosis and notification of pesticide poisonings. BMC Med Inform Decis Mak 2019 Mar 11;19(1):40 [FREE Full text] [doi: 10.1186/s12911-019-0791-2] [Medline: 30857525]
31. Rath B, Maltezou HC, Papaevangelou V, Papagrigoriou-Theodoridou MA, Alchikh M, Myles P, PEDSIDEA Network. Partnering for enhanced digital surveillance of influenza-like disease and the effect of antivirals and vaccines (PEDSIDEA). Influenza Other Respir Viruses 2019 Jul 06;13(4):309-318 [FREE Full text] [doi: 10.1111/irv.12645] [Medline: 31169347]

32. Vazquez-Lago JM, Lopez-Vazquez P, López-Durán A, Taracido-Trunk M, Figueiras A. Attitudes of primary care physicians to the prescribing of antibiotics and antimicrobial resistance: a qualitative study from Spain. Fam Pract 2012 Jun 19;29(3):352-360 [FREE Full text] [doi: 10.1093/fampra/cmr084] [Medline: 22016323]
33. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
34. Yoon CH, Ritchie SR, Duffy EJ, Thomas MG, McBride S, Read K, et al. Impact of a smartphone app on prescriber adherence to antibiotic guidelines in adult patients with community acquired pneumonia or urinary tract infections. PLoS One 2019 Jan 29;14(1):e0211157 [FREE Full text] [doi: 10.1371/journal.pone.0211157] [Medline: 30695078]
35. Wicklund E. mHealth apps with cds can help doctors make quicker, better decisions. mHealth Intelligence. URL: https://mhealthintelligence.com/news/mhealth-apps-with-cds-can-help-doctors-make-quicker-better-decisions [accessed 2020-12-29]
36. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
37. Abbasgholizadeh Rahimi S, Menear M, Robitaille H, Légaré F. Are mobile health applications useful for supporting shared decision making in diagnostic and treatment decisions? Glob Health Action 2017 Jun 25;10(sup3):1332259 [FREE Full text] [doi: 10.1080/16549716.2017.1332259] [Medline: 28838306]
38. Farzandipour M, Nabovati E, Sharif R, Arani MH, Anvari S. Patient self-management of asthma using mobile health applications: a systematic review of the functionalities and effects. Appl Clin Inform 2017 Oct;8(4):1068-1081 [FREE Full text] [doi: 10.4338/ACI-2017-07-R-0116] [Medline: 29241254]
39. Forrest G, Van Schooneveld TC, Kullar R, Schulz L, Duong P, Postelnick M. Use of electronic health records and clinical decision support systems for antimicrobial stewardship. Clin Infect Dis 2014 Oct 15;59 Suppl 3(suppl_3):S122-S133. [doi: 10.1093/cid/ciu565] [Medline: 25261539]
40. Gifford J, Vaeth E, Richards K, Siddiqui T, Gill C, Wilson L, et al. Decision support during electronic prescription to stem antibiotic overuse for acute respiratory infections: a long-term, quasi-experimental study. BMC Infect Dis 2017 Jul 31;17(1):528 [FREE Full text] [doi: 10.1186/s12879-017-2602-7] [Medline: 28760143]
41. Himes BE, Weitzman ER. Innovations in health information technologies for chronic pulmonary diseases. Respir Res 2016 Apr 05;17(1):38 [FREE Full text] [doi: 10.1186/s12931-016-0354-3] [Medline: 27048618]
42. Holstiege J, Mathes T, Pieper D. Effects of computer-aided clinical decision support systems in improving antibiotic prescribing by primary care providers: a systematic review. J Am Med Inform Assoc 2015 Jan;22(1):236-242 [FREE Full text] [doi: 10.1136/amiajnl-2014-002886] [Medline: 25125688]
43. Aungst TD. Medical applications for pharmacists using mobile devices. Ann Pharmacother 2013 Jul 02;47(7-8):1088-1095. [doi: 10.1345/aph.1s035]
44. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, Sainz-de-Abajo B, Robles M, García-Gómez JM. Mobile clinical decision support systems and applications: a literature and commercial review. J Med Syst 2014 Jan 8;38(1):4. [doi: 10.1007/s10916-013-0004-y] [Medline: 24399281]
45. Silva BMC, Rodrigues JJPC, Díez IT, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
46. Ginsburg AS, Delarosa J, Brunette W, Levari S, Sundt M, Larson C, et al. mPneumonia: development of an innovative mhealth application for diagnosing and treating childhood pneumonia and other childhood illnesses in low-resource settings. PLoS One 2015 Oct 16;10(10):e0139625 [FREE Full text] [doi: 10.1371/journal.pone.0139625] [Medline: 26474321]
47. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
48. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
49. Litvin CB, Ornstein SM, Wessell AM, Nemeth LS, Nietert PJ. Adoption of a clinical decision support system to promote judicious use of antibiotics for acute respiratory infections in primary care. Int J Med Inform 2012 Aug;81(8):521-526. [doi: 10.1016/j.ijmedinf.2012.03.002] [Medline: 22483528]
50. Panesar P, Jones A, Aldous A, Kranzer K, Halpin E, Fifer H, et al. Attitudes and behaviours to antimicrobial prescribing following introduction of a smartphone app. PLoS One 2016 Apr 25;11(4):e0154202 [FREE Full text] [doi: 10.1371/journal.pone.0154202] [Medline: 27111775]
51. Rawson T, Moore L, Hernandez B, Charani E, Castro-Sanchez E, Herrero P, et al. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately? Clin Microbiol Infect 2017 Aug;23(8):524-532 [FREE Full text] [doi: 10.1016/j.cmi.2017.02.028] [Medline: 28268133]
52. Charani E, Gharbi M, Moore L, Castro-Sanchéz E, Lawson W, Gilchrist M, et al. Effect of adding a mobile health intervention to a multimodal antimicrobial stewardship programme across three teaching hospitals: an interrupted time series study. J Antimicrob Chemother 2017 Jun 01;72(6):1825-1831 [FREE Full text] [doi: 10.1093/jac/dkx040] [Medline: 28333297]
53. Design - Material Design. Published. 2019. URL: https://material.io/design/ [accessed 2020-12-29]
54. Lindgaard G, Fernandes G, Dudek C, Brown J. Attention web designers: you have 50 milliseconds to make a good first impression!. Behaviour & Information Technology 2006 Mar;25(2):115-126. [doi: 10.1080/01449290500330448]

55. Gong J, Tarasewich P. Guidelines for Handheld Mobile Device Interface Design. In: Proceedings of DSI 2004 Annual Meeting. Presented at: Decision Sciences Institute; 2004; Boston, Massachusetts p. 3751-3756 URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.5230&rep=rep1&type=pdf [accessed 2021-01-27]
56. The Encyclopedia of Human-Computer Interaction, 2nd Ed. Interaction Design Foundation. 2014. URL: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed [accessed 2020-12-29]
57. Nayebi F, Desharnais J, Abran A. The state of the art of mobile application usability evaluation. In: 25th IEEE Canadian Conference. Inth IEEE Canadian Conference on Electrical and Computer Engineering: Vision for a Greener Future, CCECE 2012. IEEE; 2012 Presented at: 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE); 29 April-2 May 2012; Montreal, QC, Canada p. 2012-2025. [doi: 10.1109/ccece.2012.6334930]
58. Nielsen J, Budiu R. Mobile usability. In: New Riders. California, United States: New Riders Press; 2013.
59. Shneiderman B. Designing the user interface strategies for effective human-computer interaction. SIGBIO Newsl 1987 Mar;9(1):6. [doi: 10.1145/25065.950626]
60. Househ M, Borycki E, Kushniruk A. Social MediaMobile Technologies for Healthcare. Pennsylvania, United States: IGI Global; 2014.
61. Bousquet J, van Hage M. Correction to: ARIA 2016: Care pathways implementing emerging technologies for predictive medicine in rhinitis and asthma across the life cycle. Clin Transl Allergy 2019;9:53 [FREE Full text] [doi: 10.1186/s13601-019-0289-0] [Medline: 31624536]
62. Cross-platform adaptation - Material Design. Google. 2019. URL: https://material.io/design/platform-guidance/cross-platform-adaptation.html#when-to-adapt [accessed 2020-12-29]
63. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]

Abbreviations

ARIA: Allergic Rhinitis and its Impact on Asthma
CDSS: clinical decision support system
HCI: human-computer interaction
MASK: Mobile Airways Sentinel networK
mHealth: mobile health

Edited by G Eysenbach; submitted 07.04.20; peer-reviewed by B Eapen, Y Sun; comments to author 23.06.20; revised version received 30.07.20; accepted 11.11.20; published 28.01.21.

Please cite as:
Moura J, Almeida AMP, Roque F, Figueiras A, Herdeiro MT
A Mobile App to Support Clinical Diagnosis of Upper Respiratory Problems (eHealthResp): Co-Design Approach
J Med Internet Res 2021;23(1):e19194
URL: http://www.jmir.org/2021/1/e19194/
doi:10.2196/19194
PMID:33507153

Original Paper

The Role of Technology and Social Media Use in Sleep-Onset Difficulties Among Italian Adolescents: Cross-sectional Study

Nirosha Elsem Varghese[1], MSc; Eugenio Santoro[2], SCD; Alessandra Lugo[3], PhD; Juan J Madrid-Valero[4], PhD; Simone Ghislandi[1], DPhil; Aleksandra Torbica[1], PhD; Silvano Gallus[3], SCD

[1]Centre for Health and Social Care Management (CERGAS), SDA Bocconi School of Management, Bocconi University, Milan, Italy
[2]Department of Public Health, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy
[3]Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy
[4]Department of Health Psychology, Faculty of Health Science, University of Alicante, Alicante, Spain

Corresponding Author:
Eugenio Santoro, SCD
Department of Public Health
Istituto di Ricerche Farmacologiche Mario Negri IRCCS
Via Mario Negri 2
Milan, 20156
Italy
Phone: 39 02 39014562
Email: eugenio.santoro@marionegri.it

Abstract

Background: The use of technology and social media among adolescents is an increasingly prevalent phenomenon. However, there is a paucity of evidence on the relationship between frequency of use of electronic devices and social media and sleep-onset difficulties among the Italian population.

Objective: The aim of this study is to investigate the association between the use of technology and social media, including Facebook and YouTube, and sleep-onset difficulties among adolescents from Lombardy, the most populous region in Italy.

Methods: The relationship between use of technology and social media and sleep-onset difficulties was investigated. Data came from the 2013-2014 wave of the Health Behavior in School-aged Children survey, a school-based cross-sectional study conducted on 3172 adolescents aged 11 to 15 years in Northern Italy. Information was collected on difficulties in falling asleep over the last 6 months. We estimated the odds ratios (ORs) for sleep-onset difficulties and corresponding 95% CIs using logistic regression models after adjustment for major potential confounders.

Results: The percentage of adolescents with sleep-onset difficulties was 34.3% (1081/3151) overall, 29.7% (483/1625) in boys and 39.2% (598/1526) in girls. It was 30.3% (356/1176) in 11-year-olds, 36.2% (389/1074) in 13-year-olds, and 37.3% (336/901) in 15-year-olds. Sleep-onset difficulties were more frequent among adolescents with higher use of electronic devices, for general use (OR 1.50 for highest vs lowest tertile of use; 95% CI 1.21-1.85), use for playing games (OR 1.35; 95% CI 1.11-1.64), use of online social networks (OR 1.40 for always vs never or rarely; 95% CI 1.09-1.81), and YouTube (OR 2.00; 95% CI 1.50-2.66).

Conclusions: This study adds novel information about the relationship between sleep-onset difficulties and technology and social media in a representative sample of school-aged children from a geographical location that has not been included in studies of this type previously. Exposure to screen-based devices and online social media is significantly associated with adolescent sleep-onset difficulties. Interventions to create a well-coordinated parent- and school-centered strategy, thereby increasing awareness on the unfavorable effect of evolving technologies on sleep among adolescents, are needed.

(J Med Internet Res 2021;23(1):e20319) doi:10.2196/20319

KEYWORDS

sleep-onset difficulties; adolescents; social media; electronic device use

Introduction

Use of screen-based devices, such as smartphones, computer, and tablets [1], and online activities on web platforms, notably Facebook and YouTube [2,3], have grown in the recent past, particularly among adolescents. With 11 new social media users every second and each user spending more time online than before, the proportion of active social media users increased worldwide by more than 14% between January and December 2017 [4]. Specifically, among US teenagers, 95% have access to smartphones, with around 45% being online continuously [5]. The pervasive use of electronic devices has been often associated with poor sleep behaviors in adolescents. High-frequency use of social media in the hours before bed was shown to be associated with inadequate sleep [6,7]. A review of the literature found that in 90% of 67 studies conducted on the issue, the use of screen-based devices adversely affected sleep outcomes, including both sleep quantity and quality [8]. This association has been explained via three main underlying mechanisms [9], including direct time displacement from sleep duration to media use [10], increased mental, emotional, or psychological stimulation based on media content [10,11], and effects of artificial light-emitting devices on alertness, sleep health, and circadian timing [12].

More limited data are available regarding the association between social media use and sleep in the Italian population. The few studies on the issue in the United States and Northern European countries showed that the time spent on social networks, such as Facebook, as well as the time spent on the internet for playing games was significantly associated with sleep difficulties [13-15]. The share of teenagers using Facebook has decreased rapidly in the United States, with YouTube and other online social media being more commonly used [5]. The majority of studies investigating the effect of media or internet use employ measures that do not isolate the effect of online social media. Moreover, studies have rarely looked at the specific role of YouTube, the most relevant social media platform today [5]. In a recent study conducted by the Pew Research Center, 85% of the adolescents aged 13-17 years reported using YouTube, with boys reporting YouTube as their go-to platform [5]. Given the increasing prevalence of sleep problems, the continuously evolving technology, and the shift in usage preferences in social media among adolescents [16-18], the role of social media on sleep is yet to be fully elucidated.

In Italy, to our knowledge, only 2 cross-sectional studies have investigated the relationship between the use of electronic devices and sleep quality [19,20]. The first study is based on a nonrepresentative and relatively limited sample of adolescents (n=850) and finds a negative relationship between electronic device use at night and sleep quality [19]. The second and most recent study uses four waves of Health Behavior in School-aged Children (HBSC) survey data with a large sample of adolescents from 33 countries, including Italy, to show an overall negative relationship between screen time and sleep-onset difficulties [20]. The study presents an international overview of the relationship of interest, and we contribute to this literature both with additional data on use of social media and a regional focus on the issue. Secondary data from a 2008-2009 use of time survey shows that Italians sleep less compared to the European average (8 hours, 25 minutes), and the region under study, Lombardy, scores even lower on time spent sleeping (8 hours, 18 minutes). To understand the relationship between screen time and sleep-onset difficulties among the Lombards, it should be acknowledged that the temporal profile of daily activities of Italians is different from that of the rest of the world [21]. For example, using a large (n=11,000) 2017 representative sample of the Italian adolescent population, Smorti et al (2019) show very early use of smartphones (at 11 years of age), internet dependency, and long hours spent online among Italian adolescents, with females being at higher risk of dysfunctional use [22]. Findings from a 2013-2014 dated survey of 2000 children aged 9-16 years from Denmark, Italy, the United Kingdom, Romania, and Portugal show that Italian adolescents are least likely to use internet at their own home or school and most likely to use it on the go, indicating less supervised settings [23]. This also shows that Italian adolescents are more likely to use internet on their phones. Only 7% of the Italian children own or have a mobile phone that does not connect to the mobile internet, whereas these numbers range from 23% to 41% in other countries included in the sample [23]. The report shows that Denmark is the only country that has successfully integrated internet use into daily school activities, which is also a strategic site for internet awareness and safety campaigns.

Moreover, the consequences of internet dependency per se and other lifestyle habits among Italian adolescents such as unsafe sexual habits, perception of being overweight, cyberbullying, and a range of other physical and mental health problems seem to be very severe [22]. Further comparing Italy to 41 countries for adolescents aged 11-15 years shows Italy to be among the top 2 countries to report the highest proportion of adolescents reporting multiple health problems at least once in a week, positioned among the top 11 to 20 countries that have the highest proportion of adolescents skipping breakfast and among the top 5 countries to have the highest number of 15-year-old adolescents who smoke or drink alcohol at least once a week [24].

A report compiled by the World Health Organization (WHO) shows that the prevalence of recommended levels of physical activity (at least 60 minutes per day) among adolescents (aged 11-15 years) is also lower in Italy (11%) compared to several countries including Denmark (16%, 14%, and 11% for 11-, 13-, and 15-year-olds, respectively), Estonia (16% for 11- to 15-year-olds), Latvia (19% for 11- to 15-year-olds), Poland (24% for 11- to 15-year-olds), Luxembourg (30% for ages 11-14 years) and Finland (45% for ages 10-11 years and 19% for ages 14-15 years) [25]. Using a 2010-2014 representative sample of Italian individuals, it was found that smartphone use reduces quality of face-to-face social interactions, with Lombardy displaying both an increasing smartphone penetration and the lowest fraction of people reporting to see their friends at least once a week [26].

Based on the previous considerations, our study aims to understand the relationship between frequency of use of technology (including computer, tablet, and smartphone) and social media, and sleep-onset difficulties among adolescents in a large sample from a geographical location that has not been included in studies of this type previously.

Methods

Sample

The HBSC study population includes school-going children aged 11, 13, and 15 years, marking the onset of adolescence. For this analysis, we used data from the HBSC wave of 2013-2014 from Lombardy. This is the most populated Italian region, including, in 2014, 10.0 million inhabitants (ie, 17% of the 60.5 million Italian people) and 0.3 million adolescents aged 11-15 years (ie, 17% of the overall population of Italian adolescents aged 11-15 years) [27]. Within the Italian HBSC, in Lombardy schools were oversampled to obtain sufficient statistical power to derive precise frequency estimates at a regional level [28]. In this specific region, a set of questions on technology and social media use was added to the standard national questionnaire to gain in-depth knowledge in this area of interest. The HBSC primarily employs cluster sampling to choose the subjects. Sampling is done in three stages, with the first being a random selection of schools from the database of all public schools in Lombardy. Within each of the 142 participating schools, a total of 235 classes were selected, with a response rate of 89.4% on average yielding a total of 210 participating classes [28]. The classes were stratified on the basis of age with the purpose of ensuring a geographical coverage that represents the actual distribution of the population of 11-, 13-, and 15-year-olds in the Lombardy region. Design weights were applied to adjust for differences in sampling frequencies, such as cases where students in certain schools were more likely to be included in the survey. Finally, all students from the selected sample of classes were invited to complete the questionnaire. Additionally, poststratification weighting of the sample was used to ensure representativeness of pupils with respect to characteristics including school denominations, school urban-rural classification, and equal representation of boys and girls. This ensures the generalizability of the results. Details on the sampling methodology of HBSC surveys have been previously described elsewhere [29]. In Lombardy, a total of 3172 adolescents aged 11, 13, and 15 years filled a self-completion questionnaire in their schools, administered in the classrooms by teachers.

The Italian study protocol was approved by the ethics committee of University of Turin. Participation in the survey was voluntary, anonymous, and with no provision of additional benefits.

Written or oral informed consent was obtained from a parent or guardian for participants younger than 16 years.

Measures

The HBSC is an international school-based survey on adolescent health, administered every 4 years in several North American and European countries, including Italy, as a WHO collaborative cross-national study [30,31]. The survey is based on a standardized research protocol containing a theoretical framework for all the procedures, including selection of survey topics, data collection, and analysis with the objective of securing cross-national comparable data [32]. The survey instruments consist of three sets of questions: core questions similar for all participating countries to create the international dataset, optional questions on specific topics, and country-specific questions of national importance [27,31,32].

Besides general information on socioeconomic characteristics of their family, adolescents were asked to provide information on their own anthropometric variables (weight and height) and selected lifestyle habits, including tobacco smoking and alcohol drinking. The questionnaire also included specific sections on well-being, health behaviors, and social context.

Outcome

Difficulty to fall asleep was assessed from the question “over the last 6 months, how many times did you have difficulties in falling asleep?” The responses provided were (1) every day, (2) more than once a week, (3) about once a week, (4) once a month, and (5) rarely or never. For the analysis, we dichotomized the variable into two categories: participants with no sleep-onset difficulties (ie, adolescents with difficulties in falling asleep either once a month, rarely, or never) and participants with sleep-onset difficulties (ie, adolescents with difficulties in falling asleep either every day, more than once a week, or about once a week). This question and its cutoff have already been used in previous studies with the same sample [33]. Difficulty falling asleep is one of the four items of the HBSC symptom checklist used to measure psychological health and was shown to have high external and internal construct validity in a sample of Canadian adolescents [34]. The question about sleep-onset difficulties was also used to disentangle adolescents with severe sleep-onset difficulties (difficulty falling asleep every day) from adolescents with less severe sleep-onset difficulties (more than once a week, once a week, once a month, rarely or never).

Exposures

We consider four independent variables in our main analysis. Students were asked to report (1) the usage frequency of electronic devices including computers, tablets (iPad types), and smartphones for purposes such as doing homework, sending emails, chatting, tweeting, staying on Facebook, or surfing online during their free time; and (2) the use of electronic devices such as computers, consoles, tablets (iPad type), smartphones, or other devices for playing during their spare time. For both questions, the response scale provided in the questionnaire had 9 categories reported in hours per day, namely (a) never, (b) half an hour, (c) 1 hour, (d) 2 hours, (e) 3 hours, (f) 4 hours, (g) 5 hours, (h) 6 hours, and (i) 7 hours or more per day. The responses were reported for week and weekend days separately and have been shown to have considerable test-retest reliability [35,36] and acceptable criterion validity [37]. Following previous studies using the same data, we recode them to continuous measures and construct a weighted average mean, (5 × weekday + 2 × weekend day)/7 days [38]. From this weighted average, tertile categories of screen time were constructed based on the distribution of the sample. Several studies have used tertiles of screen time in similar analyses [39,40].

Students were asked to respond to how often they go on (3) social networking sites and (4) YouTube when they were connected to the internet. For both questions, the response categories provided were (1) always, (2) often, (3) sometimes, and (4) never. Likert scales of similar groupings are commonly used to indicate frequency of use of online social networking sites in several studies [41,42]. For the analysis, based on the distribution of responses, we transformed these variables into three categories, namely always, often and sometimes, or never. Information on use of social networking sites and YouTube were not collected from 11-year-olds, as social media use is restricted for children younger than 13 years in Italy as a result of Children's Online Privacy Protection Act of 1998 (COPPA). Body mass index (BMI; calculated from self-reported weight and height in kg/m^2) was categorized by considering the age- and sex-specific cutoff points adapted from Cole et al (2000) [43]. Validity of self-reported weight and height using the same data tested in several adolescent populations shows that there is a small underestimation of weight with very little consequence on overall results [44,45]. The students were also asked the number of times they smoked a cigarette and drank alcohol in a month. Studies using the HBSC dataset either use a dichotomous categorization of alcohol and smoking frequency [46] or retain the original variable as such [47]. However, studies also highlight the highly skewed nature of these two variables, thereby suggesting a dichotomous categorization of alcohol and smoking frequency into never and ever (at least one time in a month) [46].

Statistical Analysis

Odds ratios (ORs) of sleep-onset difficulties (and severe sleep-onset difficulties) and corresponding 95% CIs were estimated by multiple logistic regression models, after adjustment for a number of potential confounders based on previous literature. The model was run with all confounders included simultaneously in the regression. Considered covariates included categories of age, sex, mother's level of education (primary/secondary; high school/university; do not know), father's level of education (primary/secondary; high school/university; do not know), BMI (underweight/normal weight; obese/overweight), tobacco smoking (never; ever) and alcohol drinking (never; ever). Further, interaction tests were also performed between use of electronic devices (and social media), and sex and age. Statistical analyses were performed in SAS (SAS Institute, Cary, NC) and Stata 15 (StataCorp, College Station, TX).

Sensitivity Analysis

We do a range of sensitivity analyses. Previous studies indicate that screen time may differ between school days and weekends [48]. Hence, we construct a total measure of screen time (total time spent on electronic devices for general purposes and playing games) for weekdays and weekends separately and construct tertiles of total screen time from this unweighted measure of total screen time. For the second sensitivity analysis, we include school effects in the statistical models. It could be that going to certain schools might affect sleep outcomes differently for reasons including higher academic pressure or other school-related characteristics [49,50]. Finally, we test for different definitions of exposure to screen time and sleep-onset difficulties based on international health guidelines. We categorize frequency of overall use of electronic devices (total of time spent on electronic devices for general purposes and playing games) into exceeding 2 hours of daily screen exposure or not [51], and sleep-onset difficulties are defined as having difficulties in falling asleep every day or more than once a week [20,52]. Finally, family income is a potential confounder affecting the relationship between use of technology (and social media) and sleep problems [53]. However, given the unavailability of information on family income, we repeated our main analysis controlling for the perceived economic status of the family. The responses are categorized as very good, quite good, average, bad, and very bad, with 43 missing observations. Given the very few observations for the "very bad" category, we collapsed responses for bad and very bad into one category in our analysis.

Results

Out of 3172 adolescents, 3151 (99.3%) had available information on sleep-onset difficulties. The percentage estimates of adolescents with sleep-onset difficulties were 34.3% (1081/3151) of the total sample and 30.3% (356/1176), 36.2% (389/1074), and 37.3% (336/901) for 11-year-old, 13-year-old, and 15-year-old participants, respectively (P=.001). Female participants presented higher levels of sleep-onset difficulties (598/1526, 39.2%) than male participants (483/1625, 29.7%; P<.001).

Multivariate ORs for the relationship between adolescent sleep-onset difficulties and use of selected technologies and social media are presented in Table 1.

Table 1. Distribution of 3172 adolescents with sleep difficulties overall and by technology and social media use, with corresponding odds ratios and 95% CIs for the full sample and sex-specific samples (Lombardy, 2014).

Technology use	Total			Males			Females		
	n[a]	Difficulty falling asleep		n	Difficulty falling asleep		n	Difficulty falling asleep	
		n (%)	OR[b] (95% CI)		n (%)	OR (95% CI)		n (%)	OR (95% CI)
Electronic device use for general purposes									
1st tertile (<0.9 hours/day)	1010	292 (28.9)	1.00[c]	557	148 (26.6)	1.00[c]	453	144 (31.8)	1.00[c]
2nd tertile (0.9-2.1 hours/day)	1018	332 (32.6)	1.15 (0.94-1.42)	544	134 (24.6)	0.89 (0.66-1.19)	474	198 (41.8)	1.46 (1.08-1.97)
3rd tertile (≥2.2 hours/day)	1059	437 (41.3)	1.50 (1.21-1.85)	483	189 (39.1)	1.74 (1.30-2.32)	576	248 (43.1)	1.30 (0.96-1.77)
P value for trend	N/A[d]	N/A	<.001	N/A	N/A	<.001	N/A	N/A	.09
Electronic device use for playing games									
1st tertile (<0.8 hours/day)	1149	376 (32.7)	1.00[c]	406	109 (26.9)	1.00[c]	743	267 (35.9)	1.00[c]
2nd tertile (0.8-1.7 hours/day)	853	286 (33.5)	1.21 (0.98-1.49)	519	149 (28.7)	1.13 (0.83-1.55)	334	137 (41.0)	1.24 (0.93-1.67)
3rd tertile (≥1.8 hours/day)	1093	401 (36.7)	1.35 (1.11-1.64)	665	214 (32.2)	1.25 (0.92-1.68)	428	187 (43.7)	1.46 (1.11-1.91)
P value for trend	N/A	N/A	.003	N/A	N/A	.15	N/A	N/A	.007
Use of social networking sites									
Never/rarely	701	224 (32.0)	1.00[c]	397	104 (26.2)	1.00[c]	304	120 (39.5)	1.00[c]
Often	650	240 (36.9)	1.16 (0.91-1.48)	345	109 (31.6)	1.30 (0.92-1.84)	305	131 (43.0)	1.04 (0.74-1.47)
Always	599	256 (42.7)	1.40 (1.09-1.81)	255	89 (34.9)	1.69 (1.17-2.45)	344	167 (48.6)	1.23 (0.87-1.73)
P value for trend	N/A	N/A	.008	N/A	N/A	.005	N/A	N/A	.25
Use of YouTube									
Never/rarely	412	113 (27.4)	1.00[c]	209	44 (21.1)	1.00[c]	203	69 (34.0)	1.00[c]
Often	854	312 (36.5)	1.55 (1.18-2.04)	434	127 (29.3)	1.41 (0.94-2.13)	420	185 (44.1)	1.66 (1.14-2.42)
Always	683	294 (43.1)	2.00 (1.50-2.66)	354	130 (36.7)	2.04 (1.35-3.10)	329	164 (49.8)	1.91 (1.29-2.84)
P value for trend	N/A	N/A	<.001	N/A	N/A	.001	N/A	N/A	.001

[a] The sum does not add up to the total because of some missing values and exclusion of age 11 for social media measures.

[b] OR: odds ratio. All ORs were estimated using unconditional multiple logistic regression models after adjustment for age and sex of the child, mothers' and fathers' highest level of education, tobacco and alcohol use among adolescents, and BMI.

[c] Reference category.

[d] N/A: not applicable.

Sleep-onset difficulties were more common among adolescents with frequent use of electronic devices (OR 1.50 for ≥2.2 vs <0.9 hours/day; 95% CI 1.21-1.85; *P*<.001), use of electronic devices for playing games (OR 1.35 for ≥1.8 vs <0.8 hours/day; 95% CI 1.11-1.64; *P*=.003), online social network use (OR 1.40 for always vs never or rarely; 95% CI 1.09-1.81; *P*=.008), and YouTube use (OR 2.00 for always vs never or rarely; 95% CI 1.50-2.66; *P*<.001). Table 1 also shows the sex-specific ORs for the relationship between adolescent sleep quality and the use of selected technologies and social media. OR for sleep-onset difficulties associated with the use of YouTube was consistently large and significant in both sexes (OR 2.04 for always vs never or rarely; 95% CI 1.35-3.10; *P*=.001 for males and OR 1.91; 95% CI 1.29-2.84; *P*=.001 for females). When considering electronic device use for general purposes and use of social networking sites, the OR was significant only for males (OR 1.74 for high vs low use; 95% CI 1.30-2.32; *P*<.001 and OR 1.69; 95% CI 1.17-2.45; *P*=.005, respectively), whereas when considering use of electronic device for playing games, the OR was significant only for females (OR 1.46; 95% CI 1.08-1.97; *P*=.007).

Table 2 shows the multivariate ORs for the relationship between sleep-onset difficulties and use of selected technologies and social media stratified by age. Sleep-onset difficulties were more common among 11-year-old school-aged children with frequent use of electronic devices for general purposes (OR 1.48 for ≥2.2 vs <0.9 hours/day; 95% CI 1.03-2.13; *P*=.03) and electronic devices for playing games (OR 1.55 for ≥1.8 vs <0.8 hours/day; 95% CI 1.10-2.18; *P*=.01). Among 13-year-old adolescents, sleep-onset difficulties were more frequent among those with the use of electronic devices for general purposes (OR 1.85 for ≥2.2 vs <0.9 hours/day; 95% CI 1.29-2.66; *P*<.001) and YouTube (OR 1.88 for always vs never or rarely; 95% CI

1.26-2.81; *P*=.002). Finally, sleep-onset difficulties were more common among 15-year-old adolescents with frequent use of social networking sites (OR 1.58 for always vs never or rarely; 95% CI 1.08-1.78; *P*=.02) and use of YouTube (OR 2.10 for always vs never or rarely; 95% CI 1.39-3.18; *P*<.001).

Table 2. Distribution of 3172 adolescents with sleep difficulties overall and by technology and social media use, with corresponding odds ratios and 95% CIs for age-specific samples (Lombardy, 2014).

Technology use	11-year-old adolescents			13-year-old adolescents			15-year-old adolescents		
	n[a]	Difficulty falling asleep		n	Difficulty falling asleep		n	Difficulty falling asleep	
		n (%)	OR[b] (95% CI)		n (%)	OR (95% CI)		n (%)	OR (95% CI)
Electronic device use for general purpose									
1st tertile (<0.9 hours/day)	569	163 (28.7)	1.00[c]	264	71 (26.9)	1.00[c]	177	58 (32.8)	1.00[c]
2nd tertile (0.9-2.1 hours/day)	334	95 (28.4)	1.02 (0.73-1.42)	370	129 (34.9)	1.41 (0.98-2.05)	314	108 (34.4)	0.98 (0.65-1.49)
3rd tertile (≥2.2 hours/day)	244	89 (36.5)	1.48 (1.03-2.13)	417	184 (44.1)	1.85 (1.29-2.66)	398	164 (41.2)	1.15 (0.77-1.72)
P value for trend	N/A[d]	N/A	.03	N/A	N/A	<.001	N/A	N/A	.51
Electronic device use for playing games									
1st tertile (<0.8 hours/day)	468	129 (27.6)	1.00[c]	315	110 (34.9)	1.00[c]	366	137 (37.4)	1.00[c]
2nd tertile (0.8-1.7 hours/day)	328	98 (29.9)	1.09 (0.76-1.56)	292	103 (35.3)	1.34 (0.93-1.94)	233	85 (36.5)	1.21 (0.83-1.76)
3rd tertile (≥1.8 hours/day)	351	121 (34.5)	1.55 (1.10-2.18)	447	171 (38.3)	1.26 (0.90-1.77)	295	109 (37.0)	1.28 (0.90-1.83)
P value for trend	N/A	N/A	.01	N/A	N/A	.19	N/A	N/A	.17
Use of social networking sites (Facebook)									
Never/rarely	—[e]	—	—	440	145 (33.0)	1.00[c]	261	79 (30.3)	1.00[c]
Often	—	—	—	323	121 (37.5)	1.13 (082-1.57)	327	119 (36.4)	1.22 (0.84-1.78)
Always	—	—	—	291	120 (41.2)	1.30 (0.93-1.83)	308	136 (44.2)	1.58 (1.08-2.32)
P value for trend	—	—	—	N/A	N/A	.13	N/A	N/A	.02
Use of YouTube									
Never/rarely	—	—	—	199	54 (27.1)	1.00[c]	213	59 (27.7)	1.00[c]
Often	—	—	—	451	164 (36.4)	1.55 (1.04-2.31)	403	148 (36.7)	1.50 (1.02-2.21)
Always	—	—	—	404	168 (41.6)	1.88 (1.26-2.81)	279	126 (45.2)	2.10 (1.39-3.18)
P value for trend	—	—	—	N/A	N/A	.002	N/A	N/A	<.001

[a] The sum does not add up to the total because of some missing values and exclusion of age 11 for social media measures.

[b] OR: odds ratio. All ORs were estimated using unconditional multiple logistic regression models after adjustment for age and sex of the child, mothers' and fathers' highest level of education, tobacco and alcohol use among adolescents, and BMI.

[c] Reference category.

[d] N/A: not applicable.

[e] Not available.

In the analysis of severe sleep-onset difficulties among adolescents, l0.3% (325/3151) of adolescents had severe sleep-onset difficulties; this percentage was 11.1% (130/1176) among 11-year-olds, 10.7% (115/1074) among 13-year-olds, and 8.9% (80/901) among 15-year-olds (*P*=.24). Severe sleep-onset difficulties were more common among adolescents with frequent use of electronic devices for general purposes (OR 1.53 for ≥2.2 vs <0.9 hours/day; 95% CI 1.10-2.13; *P*=.01), electronic devices for playing games (OR 1.73 for ≥1.8 vs <0.8 hours/day; 95% CI 1.16-2.58; *P*=.001), online social networks (OR 1.73 for always vs never or rarely; 95% CI 1.16-2.58; *P*=.007), and YouTube (OR 3.18 for always vs never or rarely; 95% CI 1.89-5.34; *P*<.001) (Table 3).

Table 3. Distribution of 3172 adolescents with severe sleep difficulties overall and by technology and social media use and corresponding odds ratios and 95% CIs (Lombardy, 2014).

Technology use	n[a]	Difficulty falling asleep	
		n (%)	OR[b] (95% CI)
Electronic device use for general purpose			
1st tertile (<0.9 hours/day)	1010	88 (8.7)	1.00[c]
2nd tertile (0.9-2.1 hours/day)	1018	94 (9.2)	1.08 (0.77-1.51)
3rd tertile (≥2.2 hours/day)	1059	138 (13.0)	1.53 (1.10-2.13)
P value for trend	N/A[d]	N/A	.01
Electronic device use for playing games			
1st tertile (<0.8 hours/day)	1149	106 (9.2)	1.00[c]
2nd tertile (0.8-1.7 hours/day)	853	77 (9.0)	1.11 (0.78-1.57)
3rd tertile (≥1.8 hours/day)	1093	138 (12.6)	1.73 (1.27-2.35)
P value for trend	N/A	N/A	.001
Use of social networking sites (Facebook)			
Never/rarely	701	56 (8.0)	1.00[c]
Often	650	56 (8.6)	1.01 (0.66-1.54)
Always	599	83 (13.9)	1.73 (1.16-2.58)
P value for trend	N/A	N/A	.007
Use of YouTube			
Never/rarely	412	22 (5.3)	1.00[c]
Often	854	68 (8.0)	1.42 (0.83-2.43)
Always	683	105 (15.4)	3.18 (1.89-5.34)
P value for trend	N/A	N/A	<.001

[a] The sum does not add up to the total because of some missing values.

[b] OR: odds ratio. All ORs were estimated using unconditional multiple logistic regression models after adjustment for age and sex of the child, mothers' and fathers' highest level of education, tobacco and alcohol use among adolescents, and BMI.

[c] Reference category.

[d] N/A: not applicable.

Finally, we did not find any statistically significant interaction between use of electronic devices (and social media) and sex and age (results not shown). Tables S1, S2, and S3 in Multimedia Appendix 1 report the results of three sensitivity analyses, respectively: analysis of screen exposure by weekday and weekend (Table S1), including school effects in the statistical models (Table S2) and using a different definition of screen exposure and sleep-onset difficulties (Table S3). In all cases similar results were found in comparison with the previous analyses. Our results remain robust after controlling for perceived economic status (Table S4). Moreover, respondents perceiving worse economic status also reported higher sleep-onset difficulties (results not shown).

Discussion

Principal Findings

Our data confirms that exposure to screen-based devices, online social networking sites, and video-sharing platforms is significantly associated with difficulties in falling asleep among adolescents. Our study adds to the literature by extending the findings to Lombardy, a geographical region that has not been included in studies of this type previously. Moreover, for the first time we show that use of YouTube, an online video-sharing platform, is related to sleep-onset difficulties among adolescents.

Our findings are in broad agreement with those from a meta-analysis of 20 cross-sectional studies (N=125,198) showing 53% higher odds of poor sleep quality among those with consistent bedtime media use [54]. Previous literature also shows that the use of computers and smartphones among adolescents was associated with daytime sleep-related behaviors such as sleepiness and fatigue [55], shorter sleep duration, later bedtime, and unfavorable changes in sleep habits over time [56]. Use of PC in particular has been shown to be associated with reduced sleep duration, poor sleep, and sleep-onset latency [1,57,58]. Smartphones are easy to carry around and easier to take to bed, making it an imperative tool for adolescents. Both overall and late-night cellphone use among adolescents in several countries has been associated with several sleep difficulties [59,60]. Use of cellphones, particularly for nighttime texting, was associated with insufficient sleep [61]. In Italy, over 72% of adolescents aged 11-17 years access the internet through smartphones, with talks on a potential “no mobile phone phobia” law being drafted, particularly aiming at the younger generation [62].

Given that adolescents use not just one device, but multiple devices at the same time, our results confirm previous findings on the relationship between multiple device use and sleep problems [1]. Our results are in line with the recent findings from the international study on use of technology and sleep-onset difficulties among adolescents using the same data and variables and extend the findings to a regional level [20]. Additionally, we also confirm an exposure-response relationship between the use of electronic devices and sleep difficulties [1,17].

Given the multifunctionality of electronic devices, such as for doing homework or social networking, it is important to have specific information on the kinds of activities carried out on these platforms. Information on frequency of use of each individual social networking site is relatively less available and more valuable [6]. Information on the kind of social media used also signals the nature of the activity, such as following vloggers on YouTube or connecting with peers on Facebook. We find that use of online social media and use of YouTube among adolescents are associated with higher odds of sleep-onset difficulties. A cross-sectional survey based on 467 Scottish adolescents (11-17 years old) found that overall social media use, particularly at night, was associated with poor sleep quality [63]. In addition to sleep-onset difficulties as an outcome, various studies use standardized measurements of sleep, including the Pittsburgh Sleep Quality Index, to find significant associations with social media use in general or specific sites such as Facebook [64,65]. Using data from Millennium Cohort Study, Kelly et al (2019) found significant associations between higher frequency of weekday social media use (reported in hours) and sleep problems including inadequate sleep, sleep latency, and sleep disruption [53]. Prior studies showed that the adverse effect of social media use on sleep is mediated by cognitive activation [66] and through a behavioral effect of fear of missing out leading to delayed sleep onset [67]. Finally, our findings suggest a dose-response effect of online social media use on sleep-onset difficulties. These findings are similar to those from a recent cross-sectional study on Canadian adolescents linking cumulative effect of social media use on reduced sleep duration [14].

The use of YouTube has been steadily growing every year since its launch in 2005, and it became the most popular platform among young adults in the United States in 2018 [5]. To our knowledge, our study was the first to investigate the association between YouTube use and sleep-onset difficulties among adolescents. YouTube is a video-intensive platform with some unique characteristics including asymmetric relationships, greater self-disclosure, higher personal revelation, and developing parasocial relationships, all contributing to supplanting real relationships [68]. YouTube addiction follows the same model as internet addiction, with the frequent use of YouTube being reinforced by one-sided virtual relationships with YouTubers [68]. We found a significant association between YouTube use and sleep-onset difficulties, overall and consistently in strata of sex and age. The same relationship was also observed when considering a stricter sleep definition (ie, severe sleep difficulties).

Potential limitations of our analysis include methodological drawbacks inherent to the cross-sectional study design. First, cross-sectional designs have no dimension of time; hence, it is impossible to conclude a causal relationship. Given that reverse causality (ie, adolescents use more electronic devices and social media because of difficulties in falling asleep) could not be ruled out, our findings should be confirmed by longitudinal data.

Second, our study uses a self-reported measure of sleep latency and screen time. Self-reported sleep measures in children need to be accurate for precise identification of sleep problems [69]. Although some studies showed an overestimation of sleep latency among both adolescents and adults [70,71], self-reported sleep latency and other sleep measures obtained some consensus with objective measurements [72,73]. Our measure of sleep latency may also suffer from recall bias given the 6-month period retrospective nature of the question. However, this question has already been used and validated in different studies [20,34,74]. These types of bias could be overcome in the future by the use of apps tracking sleep quality, which are some of the most used health apps among patients and citizens [75]. In addition, in this study only sleep-onset difficulties are analyzed, whereas various aspects of sleep including sleep duration, sleep quality, other sleep disturbances, and their differences between week and weekend days should be explored in future studies. Although frequency of use of electronic devices and other social media has been self-reported, they have been shown to have considerable reliability and validity [35,37]. Self-reported measures of screen time are an inexpensive and easier means of data collection in large samples and give us detailed information on the context of the activity.

Third, we do not have data at the national level, but only for Lombardy. Although the data is geographically limited, our study population was oversampled to obtain a representative sample of adolescents aged 11, 13, and 15 years at a regional level. Moreover, Lombardy represents the most populous and richest region in Italy, accounting for one-sixth of the country’s population and one-fifth of the gross domestic product. Finally, the selection of a region from Northern Italy, which is more digitally connected than Southern Italy, fits the goal of our study.

In addition, our sample includes only young adolescents aged 11-15 years who are enrolled in schools. Although findings cannot be generalized to older adolescents, our results are crucial given that younger adolescents are at higher risk of compulsive internet use [76]. The strengths of our study include the relatively large sample size and the availability of information to control for socioeconomic characteristics and risk behavior of adolescents, thereby strengthening the internal validity of our study and the representativeness of the sample at a regional level. Moreover, to our knowledge, this is the first study to examine the relationship between screen time, social media use, and sleep-onset difficulties in a geographical location that has not been studied before in the context of this research question.

Exploiting the data on frequency of use of YouTube, a “parasocial networking site,” we also study for the first time the relationship between YouTube use and sleep-onset difficulties among adolescents.

Conclusion

Our study shows that exposure to screen-based devices and online social media is associated with adolescent sleep-onset difficulties. These findings may have important public health implications. In fact, they support interventions and guidelines to minimize excess use of technology among adolescents. In line with the recommendations proposed by several pediatric societies, including the Italian Pediatric Society, it is advised to limit the use of tablets or other electronic devices after dinner to ensure better sleep [77]. The findings also call on social media industries to be an equal player in protecting the health and well-being of young users by considering them as potential users of their services [78]. Moreover, interventions should include contributions from health professionals and educators to create a well-coordinated parent-centered strategy to increase awareness on evolving technologies, social media platforms, and the deleterious effects on sleep among adolescents.

Acknowledgments

The authors want to thank all health professionals and teachers and students who have been involved in the project and who make it possible. The HBSC Lombardy 2014 Regional Group is made up of the following people: Regional Coordinators of the HBSC study: Corrado Celata, Liliana Coppola; Regional School Office Coordinator: Bruna Baggio; HBSC Regional Research Group: Veronica Velasco, Marika Lavatelli, Giusi Gelmi, Corrado Celata; Representatives of the Local Health Units: Luca Biffi, Margherita Marella, Carlo Pellegrini, Ornella Perego, Walter De Agostini, Manuel Benedusi, Corrado Celata, Paola Duregon, Alessandra Meconi, Celeste Zagheno, Montagna Valentina Salinetti, Marina Salada, Pavia Elisabetta Mauri, Valter Drusetta, Leone Armando Lisè; Representatives of Local School Office: Antonella Giannellini, Federica Di Cosimo, Laura Peruzzo, Tullia Guerrini Rocco, Jessica Sala, Claudia Zoppi, Barbara Artioli, Laura Stampini, Maristella Colombo, Emanuela Farina, Daniela Marchesi, Linda Casalini.

This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement n° 721402. The funders had no role in the study design, execution, analyses, interpretation of the data, or decision to submit results.

Authors' Contributions

NEV drafted the initial manuscript and co-conducted the statistical analysis. ES and S Gallus conceived the original idea of the study and reviewed and revised the manuscript. AL conducted the statistical analysis and revised the study. JJMV revised and critically reviewed the manuscript for important intellectual content. S Ghislandi and AT supervised the research and critically reviewed the manuscript for important intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 30 KB - jmir_v23i1e20319_app1.docx]

References

1. Hysing M, Pallesen S, Stormark KM, Jakobsen R, Lundervold AJ, Sivertsen B. Sleep and use of electronic devices in adolescence: results from a large population-based study. BMJ Open 2015 Feb 02;5(1):e006748 [FREE Full text] [doi: 10.1136/bmjopen-2014-006748] [Medline: 25643702]
2. Gradisar M, Wolfson AR, Harvey AG, Hale L, Rosenberg R, Czeisler CA. The sleep and technology use of Americans: findings from the National Sleep Foundation's 2011 Sleep in America poll. J Clin Sleep Med 2013 Dec 15;9(12):1291-1299 [FREE Full text] [doi: 10.5664/jcsm.3272] [Medline: 24340291]
3. Uhls YT, Ellison NB, Subrahmanyam K. Benefits and Costs of Social Media in Adolescence. Pediatrics 2017 Nov;140(Suppl 2):S67-S70 [FREE Full text] [doi: 10.1542/peds.2016-1758E] [Medline: 29093035]
4. Global Digital Report. 2018. URL: https://wearesocial.com/blog/2018/01/global-digital-report-2018 [accessed 2020-10-28]
5. Anderson M, Jiang J. Teens, social media & technology 2018.: Pew Research Center; 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-10-28]
6. Shimoga SV, Erlyana E, Rebello V. Associations of Social Media Use With Physical Activity and Sleep Adequacy Among Adolescents: Cross-Sectional Survey. J Med Internet Res 2019 Jun 18;21(6):e14290 [FREE Full text] [doi: 10.2196/14290] [Medline: 31215512]

7. Reynolds AC, Meltzer LJ, Dorrian J, Centofanti SA, Biggs SN. Impact of high-frequency email and instant messaging (E/IM) interactions during the hour before bed on self-reported sleep duration and sufficiency in female Australian children and adolescents. Sleep Health 2019 Feb;5(1):64-67. [doi: 10.1016/j.sleh.2018.10.008] [Medline: 30670168]
8. Hale L, Guan S. Screen time and sleep among school-aged children and adolescents: a systematic literature review. Sleep Med Rev 2015 Jun;21:50-58. [doi: 10.1016/j.smrv.2014.07.007] [Medline: 25193149]
9. LeBourgeois MK, Hale L, Chang A, Akacem LD, Montgomery-Downs HE, Buxton OM. Digital Media and Sleep in Childhood and Adolescence. Pediatrics 2017 Nov;140(Suppl 2):S92-S96 [FREE Full text] [doi: 10.1542/peds.2016-1758J] [Medline: 29093040]
10. Cain N, Gradisar M. Electronic media use and sleep in school-aged children and adolescents: A review. Sleep Med 2010 Sep;11(8):735-742. [doi: 10.1016/j.sleep.2010.02.006] [Medline: 20673649]
11. Garrison MM, Liekweg K, Christakis DA. Media use and child sleep: the impact of content, timing, and environment. Pediatrics 2011 Jul;128(1):29-35 [FREE Full text] [doi: 10.1542/peds.2010-3304] [Medline: 21708803]
12. Chang A, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. Proc Natl Acad Sci U S A 2015 Jan 27;112(4):1232-1237 [FREE Full text] [doi: 10.1073/pnas.1418490112] [Medline: 25535358]
13. Arora T, Broglia E, Thomas GN, Taheri S. Associations between specific technologies and adolescent sleep quantity, sleep quality, and parasomnias. Sleep Med 2014 Feb;15(2):240-247. [doi: 10.1016/j.sleep.2013.08.799] [Medline: 24394730]
14. Sampasa-Kanyinga H, Hamilton HA, Chaput J. Use of social media is associated with short sleep duration in a dose-response manner in students aged 11 to 20 years. Acta Paediatr 2018 Apr;107(4):694-700. [doi: 10.1111/apa.14210] [Medline: 29363166]
15. Hawi NS, Samaha M, Griffiths MD. Internet gaming disorder in Lebanon: Relationships with age, sleep habits, and academic achievement. J Behav Addict 2018 Mar 01;7(1):70-78 [FREE Full text] [doi: 10.1556/2006.7.2018.16] [Medline: 29486571]
16. Ferrie JE, Kumari M, Salo P, Singh-Manoux A, Kivimäki M. Sleep epidemiology--a rapidly growing field. Int J Epidemiol 2011 Dec;40(6):1431-1437 [FREE Full text] [doi: 10.1093/ije/dyr203] [Medline: 22158659]
17. Twenge JM, Krizan Z, Hisler G. Decreases in self-reported sleep duration among U.S. adolescents 2009-2015 and association with new media screen time. Sleep Med 2017 Nov;39:47-53. [doi: 10.1016/j.sleep.2017.08.013] [Medline: 29157587]
18. Rowshan Ravan A, Bengtsson C, Lissner L, Lapidus L, Björkelund C. Thirty-six-year secular trends in sleep duration and sleep satisfaction, and associations with mental stress and socioeconomic factors--results of the Population Study of Women in Gothenburg, Sweden. J Sleep Res 2010 Sep;19(3):496-503 [FREE Full text] [doi: 10.1111/j.1365-2869.2009.00815.x] [Medline: 20477952]
19. Bruni O, Sette S, Fontanesi L, Baiocco R, Laghi F, Baumgartner E. Technology Use and Sleep Quality in Preadolescence and Adolescence. J Clin Sleep Med 2015 Dec 15;11(12):1433-1441 [FREE Full text] [doi: 10.5664/jcsm.5282] [Medline: 26235161]
20. Ghekiere A, Van Cauwenberg J, Vandendriessche A, Inchley J, Gaspar de Matos M, Borraccino A, et al. Trends in sleeping difficulties among European adolescents: Are these associated with physical inactivity and excessive screen time? Int J Public Health 2019 May;64(4):487-498. [doi: 10.1007/s00038-018-1188-1] [Medline: 30535677]
21. Boffi M, Colleoni M, Del Greco M. Night-time Hours and Activities of the Italians. Articulo-Journal of Urban Research 2014 Nov 11(11):articulo.3052. [doi: 10.4000/articulo.3052]
22. Smorti M, Milone A, Gonzalez Gonzalez J, Vitali Rosati G. Adolescent selfie: an Italian Society of Paediatrics survey of the lifestyle of teenagers. Ital J Pediatr 2019 May 17;45(1):62 [FREE Full text] [doi: 10.1186/s13052-019-0653-7] [Medline: 31101128]
23. Mascheroni G, Ólafsson K, Cuman A, Dinh T, Haddon L, Jørgensen H. Mobile internet access and use among European children: initial findings of the Net Children Go Mobile project. 2013. URL: http://eprints.lse.ac.uk/54244/1/Mobile%20internet%20access%20and%20use%20among%20European%20children_NCGM.pdf [accessed 2020-04-01]
24. Growing up unequal: gender and socioeconomic differences in young people's health and well-being. Health Behaviour in School-aged Children (HBSC) study: international report from the 2013/2014 survey.: World Health Organization; 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0003/303438/HSBC-No.7-Growing-up-unequal-Full-Report.pdf [accessed 2020-04-01]
25. Physical actvity country fact sheets.: World Health Organization; 2018. URL: https://www.euro.who.int/en/health-topics/disease-prevention/physical-activity/data-and-statistics/physical-activity-fact-sheets/physical-activity-country-factsheets [accessed 2020-10-28]
26. Rotondi V, Stanca L, Tomasuolo M. Connecting alone: Smartphone use, quality of social interactions and well-being. Journal of Economic Psychology 2017 Dec;63:17-26. [doi: 10.1016/j.joep.2017.09.001]
27. Istat Statistics. Resident municipal population by age, sex and marital status database on the Internet. 2014. URL: http://dati.istat.it/?lang=en# [accessed 2020-01-02]
28. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]

29. Currie C, Inchley J, Molcho M, Lenzi M, Veselska Z, Wild F. Health Behaviour in School-aged Children (HBSC) study protocol: background, methodology and mandatory items for the 2013/14 survey. St Andrews: Child and Adolescent Health Research Unit (CAHRU); 2014. URL: http://www.hbsc.org/methods/ [accessed 2020-02-01]
30. Aarø LE, Wold B, Kannas L, Rimpelä M. Health behaviour in schoolchildren. A WHO cross-national survey: A presentation of philosophy, methods and selected results of the first survey. Health Promot Int 1986;1(1):17-33. [doi: 10.1093/heapro/1.1.17]
31. Roberts C, Currie C, Samdal O, Currie D, Smith R, Maes L. Measuring the health and health behaviours of adolescents through cross-national survey research: recent developments in the Health Behaviour in School-aged Children (HBSC) study. J Public Health 2007 Apr 13;15(3):179-186. [doi: 10.1007/s10389-007-0100-x]
32. Roberts C, Freeman J, Samdal O, Schnohr CW, de Looze ME, Nic Gabhainn S, International HBSC Study Group. The Health Behaviour in School-aged Children (HBSC) study: methodological developments and current tensions. Int J Public Health 2009 Sep;54 Suppl 2:140-150 [FREE Full text] [doi: 10.1007/s00038-009-5405-9] [Medline: 19639259]
33. Kosticova M, Husarova D, Dankulincova Z. Difficulties in Getting to Sleep and their Association with Emotional and Behavioural Problems in Adolescents: Does the Sleeping Duration Influence this Association? Int J Environ Res Public Health 2020 Mar 05;17(5):1691 [FREE Full text] [doi: 10.3390/ijerph17051691] [Medline: 32150909]
34. Gariepy G, McKinnon B, Sentenac M, Elgar FJ. Validity and Reliability of a Brief Symptom Checklist to Measure Psychological Health in School-Aged Children. Child Ind Res 2015 Jul 31;9(2):471-484. [doi: 10.1007/s12187-015-9326-2]
35. Bobakova D, Hamrik Z, Badura P, Sigmundova D, Nalecz H, Kalman M. Test-retest reliability of selected physical activity and sedentary behaviour HBSC items in the Czech Republic, Slovakia and Poland. Int J Public Health 2015 Jan;60(1):59-67. [doi: 10.1007/s00038-014-0628-9] [Medline: 25471078]
36. Liu Y, Wang M, Tynjälä J, Lv Y, Villberg J, Zhang Z, et al. Test-retest reliability of selected items of Health Behaviour in School-aged Children (HBSC) survey questionnaire in Beijing, China. BMC Med Res Methodol 2010 Aug 10;10:73 [FREE Full text] [doi: 10.1186/1471-2288-10-73] [Medline: 20696078]
37. Busschaert C, De Bourdeaudhuij I, Van Holle V, Chastin SFM, Cardon G, De Cocker K. Reliability and validity of three questionnaires measuring context-specific sedentary behaviour and associated correlates in adolescents, adults and older adults. Int J Behav Nutr Phys Act 2015 Sep 17;12:117 [FREE Full text] [doi: 10.1186/s12966-015-0277-2] [Medline: 26381488]
38. Rey-López JP, Vicente-Rodriguez G, Ortega FB, Ruiz JR, Martinez-Gómez D, De Henauw S, HELENA Study Group. Sedentary patterns and media availability in European adolescents: The HELENA study. Prev Med 2010 Jul;51(1):50-55. [doi: 10.1016/j.ypmed.2010.03.013] [Medline: 20359491]
39. Gopinath B, Hardy LL, Baur LA, Burlutsky G, Mitchell P. Physical activity and sedentary behaviors and health-related quality of life in adolescents. Pediatrics 2012 Jul;130(1):e167-e174. [doi: 10.1542/peds.2011-3637] [Medline: 22689863]
40. Bucksch J, Kopcakova J, Inchley J, Troped PJ, Sudeck G, Sigmundova D, et al. Associations between perceived social and physical environmental variables and physical activity and screen time among adolescents in four European countries. Int J Public Health 2019 Jan;64(1):83-94. [doi: 10.1007/s00038-018-1172-9] [Medline: 30446772]
41. Buja A, Gallimberti L, Chindamo S, Lion C, Terraneo A, Rivera M, et al. Problematic social networking site usage and substance use by young adolescents. BMC Pediatr 2018 Nov 23;18(1):367 [FREE Full text] [doi: 10.1186/s12887-018-1316-3] [Medline: 30470215]
42. Li J, Lau JTF, Mo PKH, Su X, Wu AMS, Tang J, et al. Validation of the Social Networking Activity Intensity Scale among Junior Middle School Students in China. PLoS One 2016;11(10):e0165695 [FREE Full text] [doi: 10.1371/journal.pone.0165695] [Medline: 27798699]
43. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000 May 6;320(7244):1240-1243 [FREE Full text] [Medline: 10797032]
44. Aasvee K, Rasmussen M, Kelly C, Kurvinen E, Giacchi MV, Ahluwalia N. Validity of self-reported height and weight for estimating prevalence of overweight among Estonian adolescents: the Health Behaviour in School-aged Children study. BMC Res Notes 2015 Oct 26;8:606 [FREE Full text] [doi: 10.1186/s13104-015-1587-9] [Medline: 26502978]
45. Elgar FJ, Roberts C, Tudor-Smith C, Moore L. Validity of self-reported height and weight and predictors of bias in adolescents. J Adolesc Health 2005 Nov;37(5):371-375. [doi: 10.1016/j.jadohealth.2004.07.014] [Medline: 16227121]
46. Lee J, Choi MJ, Thornberg R, Hong JS. Exploring Sex Differences in the Association between Bullying Involvement and Alcohol and Marijuana Use among U.S. Adolescents in 6th to 10th Grade. Subst Use Misuse 2020;55(8):1203-1213. [doi: 10.1080/10826084.2020.1725054] [Medline: 32283986]
47. Horn-Hofmann C, Trost Z, Hublet A, Mrug S, Van Damme J, Vervoort T. The Relationship Between Pain Severity and Alcohol Use Among School-Aged Children and Adolescents: The Moderating Role of Drinking Motives. Pain Medicine 2018 Feb 01;19(2):284-296. [doi: 10.1093/pm/pnx016] [Medline: 28340189]
48. Sigmundová D, Sigmund E, Badura P, Vokáčová J, Trhlíková L, Bucksch J. Weekday-weekend patterns of physical activity and screen time in parents and their pre-schoolers. BMC Public Health 2016 Aug 30;16:898 [FREE Full text] [doi: 10.1186/s12889-016-3586-8] [Medline: 27576897]

49. Alsaggaf MA, Wali SO, Merdad RA, Merdad LA. Sleep quantity, quality, and insomnia symptoms of medical students during clinical years. Relationship with stress and academic performance. Saudi Med J 2016 Feb;37(2):173-182 [FREE Full text] [doi: 10.15537/smj.2016.2.14288] [Medline: 26837401]
50. Ramamoorthy S, Kamaldeen D, Ravichandran L, Sundaramahalingam M. Effect of stress on sleep hygiene among school going adolescents in Chennai. J Family Med Prim Care 2019 Sep;8(9):2917-2920 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_564_19] [Medline: 31681667]
51. American Academy of Pediatrics. Committee on Public Education. American Academy of Pediatrics: Children, adolescents, and television. Pediatrics 2001 Feb;107(2):423-426. [doi: 10.1542/peds.107.2.423] [Medline: 11158483]
52. Irish LA, Kline CE, Rothenberger SD, Krafty RT, Buysse DJ, Kravitz HM, et al. A 24-hour approach to the study of health behaviors: temporal relationships between waking health behaviors and sleep. Ann Behav Med 2014 Apr;47(2):189-197 [FREE Full text] [doi: 10.1007/s12160-013-9533-3] [Medline: 24043549]
53. Kelly Y, Zilanawala A, Booker C, Sacker A. Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study. EClinicalMedicine 2018 Dec;6:59-68 [FREE Full text] [doi: 10.1016/j.eclinm.2018.12.005] [Medline: 31193561]
54. Carter B, Rees P, Hale L, Bhattacharjee D, Paradkar MS. Association Between Portable Screen-Based Media Device Access or Use and Sleep Outcomes: A Systematic Review and Meta-analysis. JAMA Pediatr 2016 Dec 01;170(12):1202-1208 [FREE Full text] [doi: 10.1001/jamapediatrics.2016.2341] [Medline: 27802500]
55. Shochat T, Flint-Bretler O, Tzischinsky O. Sleep patterns, electronic media exposure and daytime sleep-related behaviours among Israeli adolescents. Acta Paediatr 2010 Sep;99(9):1396-1400. [doi: 10.1111/j.1651-2227.2010.01821.x] [Medline: 20377536]
56. Lemola S, Perkinson-Gloor N, Brand S, Dewald-Kaufmann JF, Grob A. Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. J Youth Adolesc 2015 Feb;44(2):405-418. [doi: 10.1007/s10964-014-0176-x] [Medline: 25204836]
57. Dorofaeff TF, Denny S. Sleep and adolescence. Do New Zealand teenagers get enough? J Paediatr Child Health 2006 Sep;42(9):515-520. [doi: 10.1111/j.1440-1754.2006.00914.x] [Medline: 16925537]
58. Punamäki R, Wallenius M, Nygård C, Saarni L, Rimpelä A. Use of information and communication technology (ICT) and perceived health in adolescence: the role of sleeping habits and waking-time tiredness. J Adolesc 2007 Aug;30(4):569-585. [doi: 10.1016/j.adolescence.2006.07.004] [Medline: 16979753]
59. Amra B, Shahsavari A, Shayan-Moghadam R, Mirheli O, Moradi-Khaniabadi B, Bazukar M, et al. The association of sleep and late-night cell phone use among adolescents. J Pediatr (Rio J) 2017;93(6):560-567 [FREE Full text] [doi: 10.1016/j.jped.2016.12.004] [Medline: 28257717]
60. Munezawa T, Kaneita Y, Osaki Y, Kanda H, Minowa M, Suzuki K, et al. The association between use of mobile phones after lights out and sleep disturbances among Japanese adolescents: a nationwide cross-sectional survey. Sleep 2011 Aug 01;34(8):1013-1020 [FREE Full text] [doi: 10.5665/SLEEP.1152] [Medline: 21804663]
61. Troxel WM, Hunter G, Scharf D. Say "GDNT": Frequency of Adolescent Texting at Night. Sleep Health 2015 Dec 01;1(4):300-303 [FREE Full text] [doi: 10.1016/j.sleh.2015.09.006] [Medline: 26981583]
62. Dipendenza da Internet, arriva la legge per curarla. Il Messaggero. 2019 Jul 22. URL: https://www.ilmessaggero.it/pay/edicola/dipendenza_internet_ultime_notizie_legge-4632455.html [accessed 2020-05-04]
63. Woods HC, Scott H. #Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. J Adolesc 2016 Aug;51:41-49. [doi: 10.1016/j.adolescence.2016.05.008] [Medline: 27294324]
64. Wolniczak I, Cáceres-DelAguila JA, Palma-Ardiles G, Arroyo KJ, Solís-Visscher R, Paredes-Yauri S, et al. Association between Facebook dependence and poor sleep quality: a study in a sample of undergraduate students in Peru. PLoS One 2013;8(3):e59087 [FREE Full text] [doi: 10.1371/journal.pone.0059087] [Medline: 23554978]
65. Levenson JC, Shensa A, Sidani JE, Colditz JB, Primack BA. The association between social media use and sleep disturbance among young adults. Prev Med 2016 Apr;85:36-41 [FREE Full text] [doi: 10.1016/j.ypmed.2016.01.001] [Medline: 26791323]
66. Harbard E, Allen NB, Trinder J, Bei B. What's Keeping Teenagers Up? Prebedtime Behaviors and Actigraphy-Assessed Sleep Over School and Vacation. J Adolesc Health 2016 Apr;58(4):426-432. [doi: 10.1016/j.jadohealth.2015.12.011] [Medline: 26874590]
67. Scott H, Woods HC. Fear of missing out and sleep: Cognitive behavioural factors in adolescents' nighttime social media use. J Adolesc 2018 Oct;68:61-65. [doi: 10.1016/j.adolescence.2018.07.009] [Medline: 30031979]
68. de Bérail P, Guillon M, Bungener C. The relations between YouTube addiction, social anxiety and parasocial relationships with YouTubers: A moderated-mediation model based on a cognitive-behavioral framework. Computers in Human Behavior 2019 Oct;99:190-204. [doi: 10.1016/j.chb.2019.05.007]
69. Erwin AM, Bashore L. Subjective Sleep Measures in Children: Self-Report. Front Pediatr 2017;5:22 [FREE Full text] [doi: 10.3389/fped.2017.00022] [Medline: 28243584]
70. Wolfson AR, Carskadon MA, Acebo C, Seifer R, Fallone G, Labyak SE, et al. Evidence for the validity of a sleep habits survey for adolescents. Sleep 2003 Mar 15;26(2):213-216. [doi: 10.1093/sleep/26.2.213] [Medline: 12683482]

71. Chambers MJ. Actigraphy and insomnia: a closer look. Part 1. Sleep 1994 Aug;17(5):405-8; discussion 408-10. [Medline: 7880249]
72. Zinkhan M, Berger K, Hense S, Nagel M, Obst A, Koch B, et al. Agreement of different methods for assessing sleep characteristics: a comparison of two actigraphs, wrist and hip placement, and self-report with polysomnography. Sleep Med 2014 Sep;15(9):1107-1114. [doi: 10.1016/j.sleep.2014.04.015] [Medline: 25018025]
73. Wetter DW, Young TB. The relation between cigarette smoking and sleep disturbance. Prev Med 1994 May;23(3):328-334. [doi: 10.1006/pmed.1994.1046] [Medline: 8078854]
74. Pallesen S, Hetland J, Sivertsen B, Samdal O, Torsheim T, Nordhus IH. Time trends in sleep-onset difficulties among Norwegian adolescents: 1983--2005. Scand J Public Health 2008 Nov;36(8):889-895. [doi: 10.1177/1403494808095953] [Medline: 19004908]
75. Mosconi P, Radrezza S, Lettieri E, Santoro E. Use of Health Apps and Wearable Devices: Survey Among Italian Associations for Patient Advocacy. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e10242. [doi: 10.2196/10242] [Medline: 30664455]
76. Espinoza G, Juvonen J. The pervasiveness, connectedness, and intrusiveness of social network site use among young adolescents. Cyberpsychol Behav Soc Netw 2011 Dec;14(12):705-709. [doi: 10.1089/cyber.2010.0492] [Medline: 21668346]
77. Un bambino su quattro soffre di disturbi del sonno. Le 10 regole d'oro per far dormire i piccoli, i più grandi (e anche mamma e papà). Società Italiana di Pediatria. URL: https://sip.it/2019/03/15/un-bambino-quattro-soffre-disturbi-del-sonno-le-10-regole-doro-far-dormire-piccoli-piu-grandi-anche-mamma-papa/ [accessed 2020-10-28]
78. Life in ‘likes’: Children’s Commissioner report into social media use among 8-12 year olds.: Children's Commissioner; 2018 Jan 04. URL: https://www.basw.co.uk/system/files/resources/basw_94916-10.pdf [accessed 2020-05-06]

Abbreviations

BMI: body mass index
HBSC: Health Behavior in School-aged Children
OR: odds ratio
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.05.20; peer-reviewed by J Piqueras, K Smith; comments to author 19.09.20; revised version received 30.09.20; accepted 28.10.20; published 21.01.21.

Please cite as:
Varghese NE, Santoro E, Lugo A, Madrid-Valero JJ, Ghislandi S, Torbica A, Gallus S
The Role of Technology and Social Media Use in Sleep-Onset Difficulties Among Italian Adolescents: Cross-sectional Study
J Med Internet Res 2021;23(1):e20319
URL: http://www.jmir.org/2021/1/e20319/
doi:10.2196/20319
PMID:33475517

Original Paper

Patterns of Routes of Administration and Drug Tampering for Nonmedical Opioid Consumption: Data Mining and Content Analysis of Reddit Discussions

Duilio Balsamo[1], MS; Paolo Bajardi[2], PhD; Alberto Salomone[3], PhD; Rossano Schifanella[2,4], PhD

[1]Department of Mathematics, University of Turin, Turin, Italy

[2]ISI Foundation, Turin, Italy

[3]Department of Chemistry, University of Turin, Turin, Italy

[4]Department of Computer Science, University of Turin, Turin, Italy

Corresponding Author:
Duilio Balsamo, MS
Department of Mathematics
University of Turin
Via Carlo Alberto, 10
Turin, 10124
Italy
Phone: 39 3331371972
Email: duilio.balsamo@unito.it

Abstract

Background: The complex unfolding of the US opioid epidemic in the last 20 years has been the subject of a large body of medical and pharmacological research, and it has sparked a multidisciplinary discussion on how to implement interventions and policies to effectively control its impact on public health.

Objective: This study leverages Reddit, a social media platform, as the primary data source to investigate the opioid crisis. We aimed to find a large cohort of Reddit users interested in discussing the use of opioids, trace the temporal evolution of their interest, and extensively characterize patterns of the nonmedical consumption of opioids, with a focus on routes of administration and drug tampering.

Methods: We used a semiautomatic information retrieval algorithm to identify subreddits discussing nonmedical opioid consumption and developed a methodology based on word embedding to find alternative colloquial and nonmedical terms referring to opioid substances, routes of administration, and drug-tampering methods. We modeled the preferences of adoption of substances and routes of administration, estimating their prevalence and temporal unfolding. Ultimately, through the evaluation of odds ratios based on co-mentions, we measured the strength of association between opioid substances, routes of administration, and drug tampering.

Results: We identified 32 subreddits discussing nonmedical opioid usage from 2014 to 2018 and observed the evolution of interest among over 86,000 Reddit users potentially involved in firsthand opioid usage. We learned the language model of opioid consumption and provided alternative vocabularies for opioid substances, routes of administration, and drug tampering. A data-driven taxonomy of nonmedical routes of administration was proposed. We modeled the temporal evolution of interest in opioid consumption by ranking the popularity of the adoption of opioid substances and routes of administration, observing relevant trends, such as the surge in synthetic opioids like fentanyl and an increasing interest in rectal administration. In addition, we measured the strength of association between drug tampering, routes of administration, and substance consumption, finding evidence of understudied abusive behaviors, like chewing fentanyl patches and dissolving buprenorphine sublingually.

Conclusions: This work investigated some important consumption-related aspects of the opioid epidemic using Reddit data. We believe that our approach may provide a novel perspective for a more comprehensive understanding of nonmedical abuse of opioids substances and inform the prevention, treatment, and control of the public health effects.

(J Med Internet Res 2021;23(1):e21212) doi:10.2196/21212

KEYWORDS

routes of administration; drug tampering; Reddit; word embedding; social media; opioid; heroin; buprenorphine; oxycodone; fentanyl

Introduction

Background

In the last decade, the United States witnessed an unprecedented growth of deaths due to opioid drugs [1], which sparked from overprescriptions of semisynthetic opioid pain medication such as oxycodone and hydromorphone and evolved in a surge of abuse of illicit opioids like heroin [2,3] and powerful synthetic opioids like fentanyl [4,5]. Alongside traditional medical, pharmacological, and public health studies on the nonmedical adoption of prescription opioids [6-14], several phenomena related to the opioid epidemic have recently been successfully tackled through a digital epidemiology [15-18] approach. Researchers have used digital and social media data to perform various tasks, including detecting drug abuse [19,20], forecasting opioid overdose [21], studying transition into drug addiction [22], predicting opioid relapse [23], and discovering previously unknown treatments for opioid addiction [24]. A few recent studies investigated the temporal unfolding of the opioid epidemic in the United States by leveraging complementary data sources different from the official US Centers for Disease Control and Prevention data [2,25-28] and using social media like Reddit [29,30].

Pharmacology research is interested in understanding the consequences of various routes of administration (ROA), that is, the paths by which a substance is taken into the body [6,31,32], due to the different effects and potential health-related risks tied to them [10,33,34]. Researchers have estimated the prevalence of routes of administration for nonmedical prescription opioids [9,31,32,35] and opiates [36,37]; however, they rarely consider less common ROA, such as rectal, transdermal, or subcutaneous administration [32,38], leaving the mapping of nonmedical and nonconventional administration behaviors greatly unexplored [39,40]. Many of these studies [31,32,35] acknowledge that drug tampering, that is, the intentional chemical or physical alteration of medications [41], is an important constituent of drug abuse. The alteration of the pharmacokinetics of opioids through drug-tampering methods, together with unconventional administration, may potentially lead to very different addictive patterns and ultimately have unexpected health-associated risks [33]. Research has also been focused on developing tamper-resistant and abuse-deterrent drug formulations. However, to the best of our knowledge, no large-scale empirical evidence has been found to unveil the relationships between substance manipulation, unconventional ROA, and nonmedical substance administration.

Goals

This paper seeks to complement current studies widening the understanding of opioid consumption patterns by using Reddit, a social content aggregation website, as the primary data source. This platform is structured into subreddits, user-generated and user-moderated communities dedicated to the discussion of specific topics (Multimedia Appendix 1). Due to fair guarantees of anonymity, no limits on the number of characters in a post, and a large variety of debated topics, this platform is often used to uninhibitedly discuss personal experiences [42]. Reddit constitutes a nonintrusive and privileged data source to study a variety of issues [43,44], including sensitive topics such as mental health [45], weight loss [46], gender issues [47], and substance abuse [22,24].

This study’s contributions are manifold. First, leveraging and expanding a recent methodology proposed by Balsamo et al [30], we identified a large cohort of opioid firsthand users (ie, Reddit users showing explicit interest in firsthand opioid consumption) and characterized their habits of substance use, administration, and drug tampering over a period of 5 years. Second, using word embeddings, we identified and cataloged a large set of terms describing practices of nonmedical opioids consumption. These terms are invaluable to performing exhaustive and at-scale analyses of user-generated content from social media, as they include colloquialisms, slang, and nonmedical terminology that is established on digital platforms and hardly used in the medical literature. We provided a longitudinal perspective on online interest in the opioids discourse and a quantitative characterization of the adoption of different ROA, with a focus on the less-studied yet emerging and relevant practices. We have made available the ROA taxonomy and the corresponding vocabulary to the research community. Third, we quantified the strength of association between ROA and drug-tampering methods to better characterize emerging practices. Finally, we investigated the interplay between the previous 3 dimensions, measuring odds ratios to shed light on the “how” and “what” facets of the opioid consumption phenomenon. We studied a wide spectrum of opioid forms, referred to as “opioids” throughout, ranging from prescription opioids to opiates and illegal opioid formulations. To the best of our knowledge, our contributions are original in both breadth and depth, outlining a detailed picture of nonmedical practices and abusive behaviors of opioid consumption through the lenses of digital data.

Methods

Data

We refer to a publicly available Reddit data set [48] that contains all the subreddits published on the platform since 2007 [44,49]. In this work, we analyzed the textual part of the submissions and the comments collected from 2014 to 2018. We preprocessed each year separately, filtering out the subreddits with less than 100 comments in a year. We used spaCy [50] to remove English stop words, inflectional endings, and tokens with less than 100 yearly appearances. We adopted a bag-of-words model, resulting in a vocabulary of different lemmas for each year. Vocabulary sizes ranged from 300,000 to 700,000 lemmas, with a size growth of approximately 30% each year. In Table 1, the number of unique comments and unique active users per year is reported. A steady growth of approximately 30% per year both in the volume of conversations and in the active user base is observed.

All the analyses in this work were performed on a subset of subreddits related to opioid consumption, which were identified using the procedure described here. For space constraints, we restricted the analyses of odds ratios to comments and submissions created during 2018. Similar to a vast body of users' activities on social media platforms [51-53], the distribution of posts per user shows a heavy tail, with the majority of users posting few comments and the remaining minority (eg, core users and subreddit moderators) producing a large portion of the content. Moreover, a nonnegligible percentage of posts, respectively 25% and 7% of submissions and comments, were produced by authors who deleted their usernames.

Table 1. Data set statistics.

Year	Reddit comments, n	Reddit authors, n	Opiates subreddits, n	Opiates comments, n	Opiates authors, n	Authors' prevalence
2014	545,720,071	8,149,234	19	386,984	12,381	0.0015
2015	699,245,245	10,673,990	19	470,609	15,888	0.0015
2016	840,575,089	12,849,603	25	612,619	21,791	0.0017
2017	1,045,425,499	14,219,062	30	866,023	28,358	0.0020
2018	1,307,123,219	18,158,464	25	919,036	33,700	0.0019

Analytical Pipeline

The methodology adopted in this paper consists of several steps. First, we identified a cohort of opioid firsthand users and the subreddits related to opioid consumption through a semiautomatic algorithm. Second, we trained a word-embedding language model to capture the latent semantic features of the discourse on the nonmedical use of opioids. Third, we exploited the embedded vectors to extend an initial set of medical terms known from the literature, (eg, opioid substance names, ROA, and drug-tampering methods) to nonmedical and colloquial expressions. The terms were organized in a taxonomy that provides a conceptual map on the topic. Moreover, we studied the temporal evolution of the popularity of the main opioid substances and ROA. Ultimately, we measured the strength of the associations between opioid substances, routes of administration, and drug-tampering techniques in 2018.

Identification of Firsthand Opioid Consumption on Reddit

We leveraged a semiautomatic information retrieval algorithm developed to identify relevant content related to a topic of interest [30] to collect opioid-related conversations on Reddit yearly. This approach aims at retrieving topic-specific documents by expressing a set of initial keywords of interest; here, it identified relevant subspaces of discussion via an iterative query expansion process, retaining a list of terms Q_y and a list of subreddits S_y ranked by relevance for each year.

We merged all the query terms in a set containing 67 terms. To ensure that the sets S_y were effectively referring to the opioid-related topics and in particular to nonmedical opioid consumption, we performed a manual inspection on the union of the top 150 subreddits for each year, for a total of 554 subreddits. Three independent annotators, including a domain expert specialized in antidoping analyses, read a random sample of 30 posts, checking for subreddits (1) mostly focused on discussing the use of opioids, (2) mostly focused on firsthand usage, and (3) not focused on medical treatments. This yielded a total of 32 selected subreddits, with a Fleiss κ interrater agreement of κ=0.731, which suggests a substantial agreement, according to Landis and Koch [54]. Multimedia Appendix 2 presents a complete list of the subreddits broken down by year.

Automatic language detection, performed with langdetect [55], cld2 [56], and cld3 [57], showed that the majority of posts (about 90%) were in English, approximately 5% were non-English messages, and the rest were too short or full of jargon and emojis to algorithmically detect any language. Assuming that an author who writes in one of the selected subreddits is personally interested in the topic, we identified a cohort of 86,445 unique opioid firsthand users involved in direct discussions of opioid usage across the period of study. Summary statistics are reported in Table 1. In particular, for each year, we computed the number of unique active users and the volume of comments shared, as well as the user's relative prevalence over the entire amount of Reddit activity. We observed growth from 2014 to 2017, ranging from 15 to 19 users interested in opioid consumption out of every 100,000 Reddit users.

Vocabulary Expansion

The methodology to extend the vocabulary on opioid-related domains with user-generated slang and colloquial forms was implemented in 2 steps. First, we trained a word-embedding model (word2vec [58]), which learns semantic relationships in the corpus during training and maps their terms to vectors in a latent vectorial space, with all the comments and submissions in our subreddit data set (relevant training parameters are displayed in Multimedia Appendix 3). Second, starting from a set of seed terms K (eg, a list of known opioid substances), we expanded the vocabulary by navigating the semantic neighborhood of each element $w \in K$ in the embedded space, considering the n=20 semantically closest elements in terms of cosine similarity. We merged the results in a candidate expansion set, , together with the seed terms K if not already included. Based on the knowledge of a domain expert (a clinical and forensic toxicologist) and with the help of search engine queries and a crowdsourced online dictionary for slang words and phrases (Urban Dictionary [59]) to understand the most unusual terms, we manually selected and categorized the relevant neighboring terms, obtaining an extended vocabulary *V*. Figure 1 shows an example of the expansion procedure in which the high-dimensional vectors are projected to 2 dimensions using the uniform manifold approximation and projection (UMAP) algorithm [60].

As a sensitivity analysis, we compared the effectiveness of an alternative embedding model (GloVe [61]) for topical coherence. In the case of vocabulary expansion of opioid substance terms, that is, using as seeds, the 2 models captured 100 terms in common out of their respective candidate terms, with word2vec showing a higher number and a larger percentage of accepted terms (Table 2). Moreover, the volume of comments that included an accepted term was almost double when using the vocabulary of word2vec rather than the vocabulary of GloVe. Hence, we chose word2vec as the reference word-embedding model.

Figure 1. Two-dimensional projection of the word2vec embedding, modeling the semantic relationships among terms in the Reddit opioids data set. Filled markers represent the seed terms K. Expansion terms, represented with hollow markers, are colored according to their respective initial term if accepted or in gray if discarded. The nature of the relationships between neighboring terms varies, representing (1) equivalence (eg, synonyms), (2) common practices (eg, the use of methadone for addiction maintenance), or (3) co-use (eg, the cluster of heroin, cocaine, and methamphetamine).

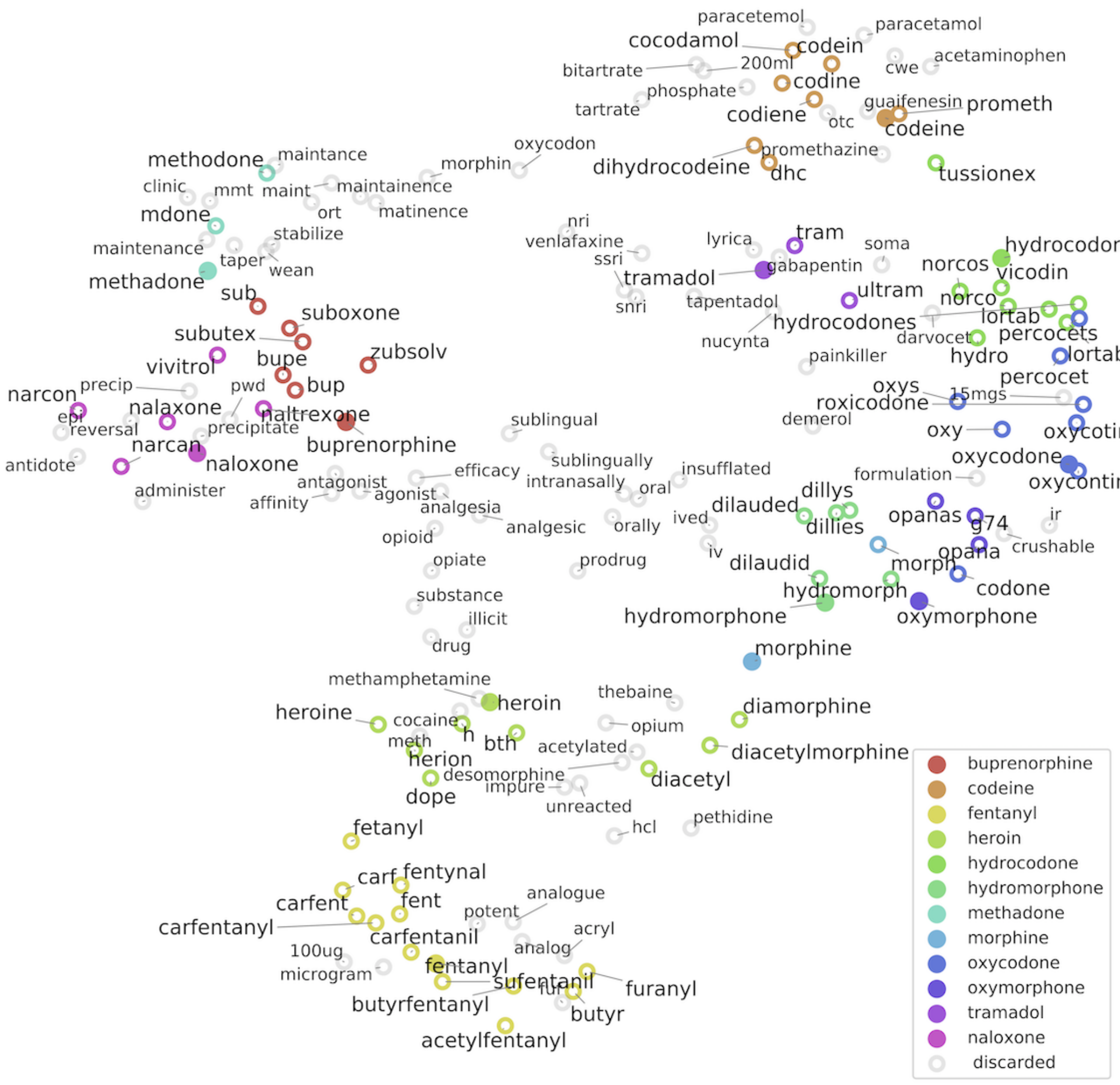


Table 2. Comparison of term expansions of opioid substances for the 2 trained models.

Model	Candidate terms, n	Accepted terms, n (%)	Comments[a], n
word2vec	297	128 (43.1)	225,165
GloVe	369	110 (29.8)	144,564

[a] Comments in the corpus mentioning at least one term of the respective accepted terms for vocabulary expansion.

Strength of Association Between Opioid Substances, ROA, and Drug Tampering

We evaluated the odds ratios (ORs) to quantify the pairwise strength of the association between substance use and ROA, substance use and drug-tampering methods, and ROA and drug-tampering methods. Under the assumption that co-mention was a proxy for associating a substance to its ROA (or drug-tampering method), we focused on the posts that contained a reference to terms in each domain, evaluating contingency tables and odds ratios. Odds ratios, significance, and confidence intervals were estimated using chi-square tests implemented in the statsmodel Python package [62], with the significance level set to α=.01. As a sensitivity analysis, we assessed the effect of the proximity of terms on the characterization of odds ratios. We modified the definition of co-occurrence, introducing a distance threshold ρ at sentence level. We explored the range $\rho \in \{0...5\}$, where ρ=0 indicates that co-occurrence appears within the same sentence, and ρ>0 measures the distance in both directions (eg, ρ=1 for the preceding and consecutive sentences). The value $\rho=\infty$ indicates the scenario in which we considered the entire post as reference. Accordingly, given a threshold ρ in the construction of the contingency table, the co-occurrence event between 2 terms is conditioned to their distance being less than or equal to ρ. Conversely, we considered terms to be separate events in cases of distance above the threshold. It is important to consider that the OR measures do not imply any form of causation but rather surface correlations that could be used in hypothesis formation. To better interpret the results of this analysis, in some cases, manual inspection of the comments mentioning the variables under investigation was performed following the directives on privacy and ethics (see the "Ethics and Privacy" section).

Results

Characterizing Interest in Opioids, ROA, and Drug-Tampering Methods

We applied the methodology described in the "Vocabulary Expansion" section to extract and expand domain-specific vocabularies and to characterize the temporal unfolding of interest in different opioid substances, routes of administration, and drug-tampering methodologies. We started from a review of the relevant medical research, collecting an initial set of terms referring to the most common opioid substances, ROA [6,10,31,34,38,39,41,63,64], and drug-tampering methods [41,63]. We expanded the original set with neighboring terms in a low-dimensional embedding space, and the outputs were reviewed and organized by a domain expert. The resulting vocabulary for opioid substances is shown in Table 3. It is worth noting that the vocabulary expansion procedure considerably increased the richness of the terminology related to the domain of interest and, consequently, the volume of conversations on Reddit that contained these terms. For example, for the heroin category, we observed a 62% growth in the retrieved relevant conversations (Table 3). We investigated the temporal unfolding of the popularity of the opioid substances, measured as the fraction of authors mentioning a substance over the entire opioid firsthand user base, for each trimester from 2014 to 2018. A binary characterization of the mentioning behavior at the user level was considered to discount potential biases due to users with high activity. We also provided a relative measure of popularity to account for the constantly increasing volume of active users on Reddit. Figure 2 shows a decrease in the usage of heroin and a rise in fentanyl and codeine.

The resulting vocabulary for routes of administration was further organized in a 2-level hierarchical structure, reported in Table 4. It is worth noting that the taxonomy does not have a strict medical interpretation, nor was it intended to be a comprehensive review. However, it can give structure to otherwise unstructured collections of words and help in the interpretation of the results.

Table 3. Vocabulary of opioid substances. Starting from a candidate expansion set *E*-, comprising 297 unique terms, the final expansion terms considered equivalent to a substance were gathered in the same class.

Substance	Terms	Δ volume, %[a]
Heroin	bth[b], diacetylmorphine, diamorphine, dope, ecp[b], goofball, goofballs, gunpowder, h, herion[b], heroin[b], heroine, heron, smack, speedball, speedballing, speedballs[b], tar	62
Buprenorphine	bup, bupe[b], buprenorphine[b], butrans[b], sub, suboxone[b], subutex[b], zub, zubsolv[b]	61
Hydrocodone	hydro, hydrocodone[b], hydrocodones[b], lortab[b], lortabs[b], norco[b], norcos[b], tuss, tussionex[b], vic, vicoden, vicodin[b], vicodins[b], vicoprofen[b], vics[b], vikes, viks, zohydro[b]	38
Codeine	cocodamol, codein[b], codeine[b], codiene[b], codine, dhc, dihydrocodeine[b], prometh, sizzurp, syrup	28
Oxymorphone	g74, opana[b], opanas, oxymorphone[b], panda	25
Tramadol	desmethyltramadol, dsmt, tram, tramadol[b], ultram[b]	22
Hydromorphone	dil, dilauded, dilaudid[b], dilaudids, dillies[b], dilly, dillys, diluadid[b], hydromorph[b], hydromorphone[b]	21
Oxycodone	15s, 30s, codone, contin, ms, oc, ocs, oxy[b], oxycodone[b], oxycontin[b], oxycontins, oxycotin[b], oxys[b], perc[b], percocet[b], percocets[b], percoset, percosets, percs[b], perk, roxi[b], roxicodone[b], roxie[b], roxies[b], roxis[b], roxy[b], roxycodone[b], roxys[b]	14
Morphine	kadian, morph, morphine[b]	5
Fentanyl	acetylfentanyl[b], butyr, butyrfentanyl, carf, carfent, carfentanil[b], carfentanyl, duragesic[b], fent[b], fentanyl[b], fents, fentynal, fetanyl, furanyl, sufentanil, u47700	4
Antagonist	nalaxone[b], naloxone[b], naltrexone, narcan[b], narcon, revia, viv, vivitrol[b]	1
Methadone	mdone, methadone[b], methodone[b]	1

[a] The increase in the volume of occurrences of a substance using the terms in the expanded vocabulary compared with only using the terms in .

[b] Term in .

Figure 2. Popularity of opioid substances among opioid firsthand users on Reddit. Each line represents the share of opioid firsthand users mentioning an opioid substance, measured quarterly from 2014 to 2018.

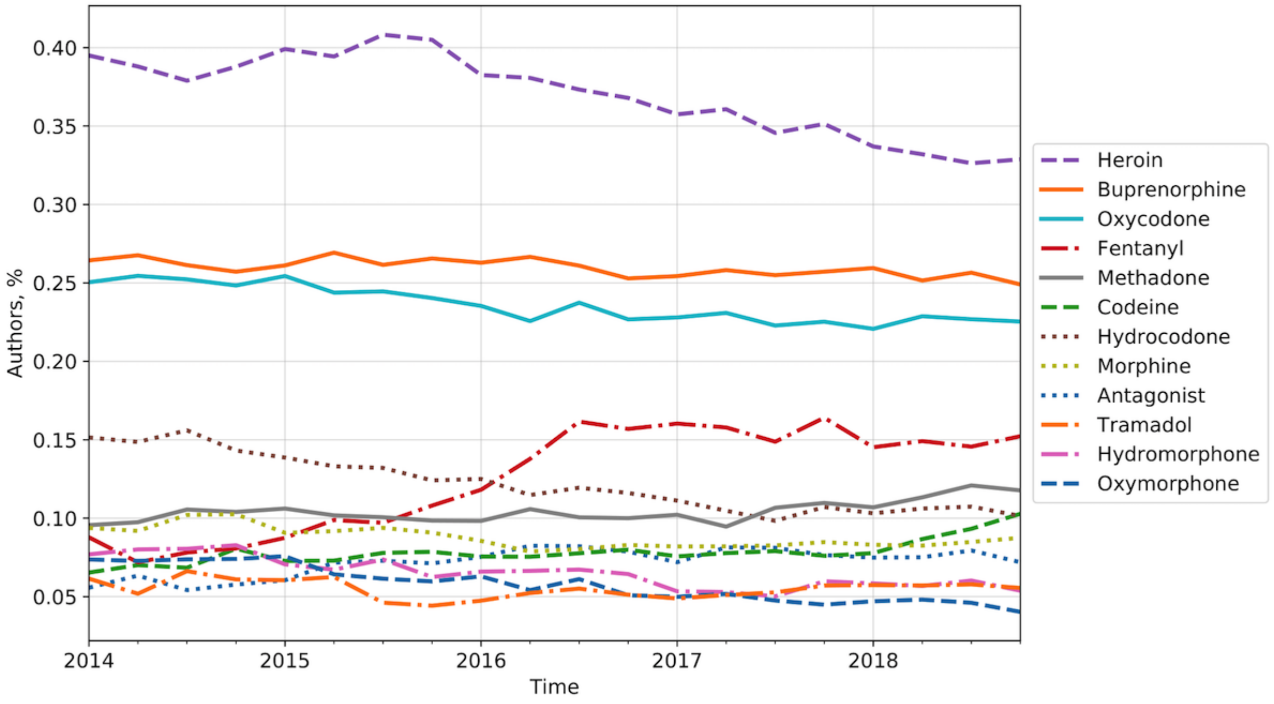


Table 4. Taxonomy defining the ROA categories and their corresponding terms. Primary ROA include all the expansion terms considered for the appropriate secondary ROA (original candidate expansion set comprised 199 unique terms).

Primary and secondary ROA[a]	Terms
Ingestion	
Oral	bolus, buccal, gulp, mouth, mouthful, oral[b], orally, swallow[b]
Sublingual	sublingual[b], sublingually, tongue, tounge
Drink	chug, drink, pour, pourin, sip[b], sipper, sippin, swig, swish
Chew	chew[b], chewy, chomp, gum
General ingestion	ingest[b], ingestion
Inhalation	
Intranasal	intranasal, intranasally, nasal, nasally, nose, nostril, rail, sniff[b], sniffer, sniffin, snoot, snooter, snort[b], snorter, tooter
General inhalation	breath, breathe, dab, exhale, inhalation, inhale[b], insufflate, insufflated, insufflating, insufflation, puff, toke, tokes, vap, vape, vaped, vapes, vaping, vapor, vaporise, vaporize, vaporizer, vapour
Smoking	bong, fume, hookah, pipe, smoke[b], smoker, smokin, spliff
Injection	
Intramuscular	deltoid, imed, iming, intramuscular[b], intramuscularly
Subcutaneous	subcutaneous[b], subcutaneously, subq
Intravenous	arterial, bloodstream, intravenous[b], intravenously, iv, ivd, ived, iving, ivs, vein, venous
General injection	bang, inject[b], injectable, injection, parenteral, shoot, shot
Rectally	anal, anally, boof, boofed, boofing, bunghole, butt, pooper, rectal[b], rectally
Other ROA	
Dermal	cutaneous, dermis, transdermal[b], transdermally
Urogenital	vaginal
Intrathecal	intrathecal

[a] ROA: routes of administration.

[b] Seed term *K*.

Figure 3 shows the estimated temporal evolution of the relative popularity of the routes of administration from 2014 to 2018, measured in quarterly snapshots. Finally, we extracted and organized the vocabulary related to drug-tampering techniques, as shown in Table 5. In this paper, we considered the act of chewing pills a second-level route of administration under the ingestion category [8,31,32] instead of a drug-tampering method, as some research might suggest [41].

Figure 3. Popularity of routes of administration among opioid firsthand users on Reddit. Each line represents the fraction of opioid firsthand users mentioning an ROA-related term, measured quarterly from 2014 to 2018. Thick lines represent the share of authors mentioning primary ROA, evaluated by aggregating the contributions of all the corresponding secondary ROA. ROA: routes of administration.

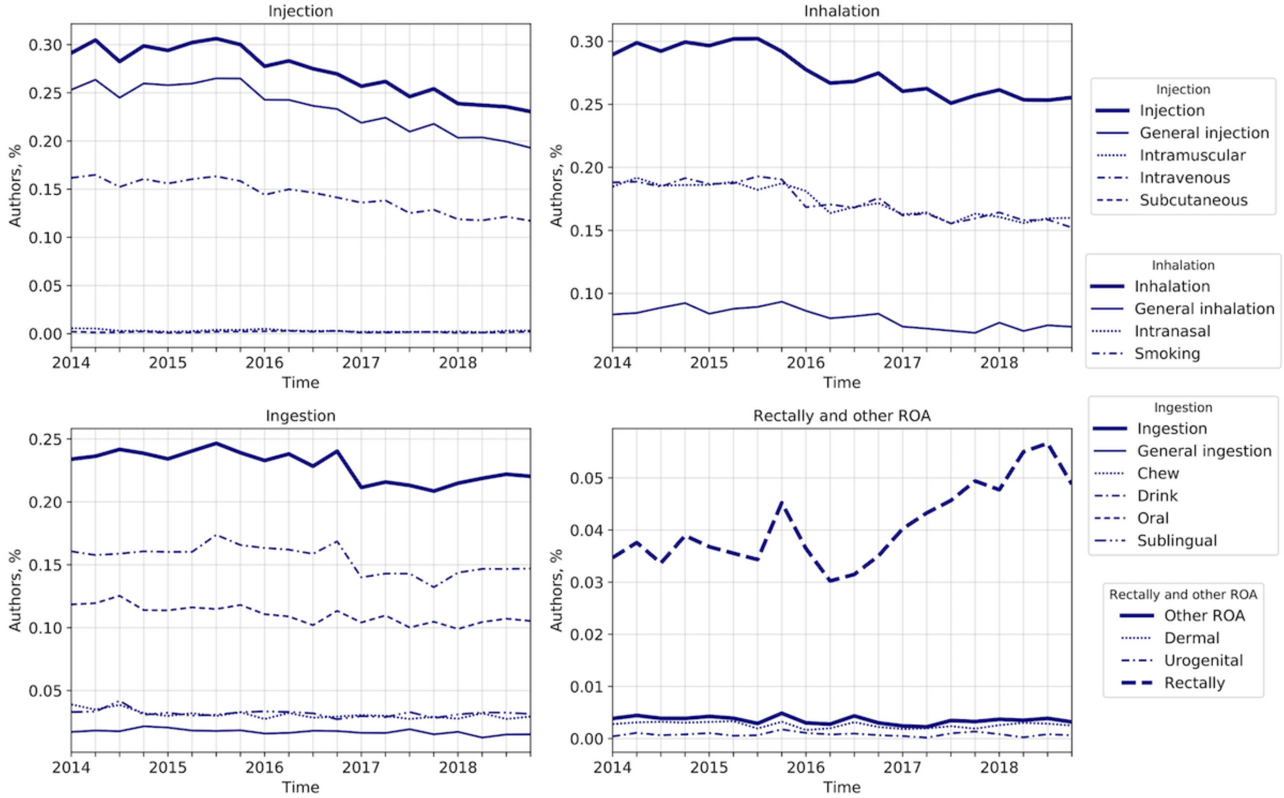


Table 5. Vocabulary of drug-tampering methods. Expansion terms referring to the same drug-tampering method are grouped in the corresponding transformation classes (original candidate expansion set comprised 179 unique terms).

Transformation	Terms
Brew	brew[a], brewer, homebrew[a]
Concentrate	concentrate[a], concentrate, concentration, purify
Dissolve	desolve, dilute, disolve, disolved, disolves, dissolve[a], solute, solution, soluble, soluable
Evaporate	evap, evaporate
Extract	cwe[a], extract[a], extract, extraction
Grind	chop, crush[a], crushable, crusher, grind[a], grinded, grinder, ground, pulverize
Heat	boil, heat[a], melt, microwave, overheat, simmer
Infusion	infuse, infusion[a], tea, tincture
Peel	peal, peel, shave
Soak	soak[a], submerge
Wash	rewash, rinse, wash

[a] Seed term *K*.

Characterizing the Associations Between Opioid Substances, ROA, and Drug Tampering

To investigate the strength of association between routes of administration, drug tampering, and opioid substances and to shed light on the interplay between the "how" and the "what" dimensions of opioid consumption, we estimated the ORs, 95% confidence intervals, *P* values, and volume of the co-mentions among substances, routes of administration, and drug-tampering methods. The number of sentences in Reddit posts vary greatly, but the posts are generally quite short (approximately 50% of them have 2 sentences or less, as seen in Multimedia Appendix 4). However, as about 20% of posts have more than 10 sentences, one should be cautious in adopting a bag-of-words approach to measure co-occurring terms. To limit the chance of including spurious correlations due to the co-occurrences of terms far apart in the posts, we conservatively selected ρ=1 (ie, considering only the co-occurrence of terms within a sentence or in the first adjacent sentences) for computing the OR. Figure 4 shows in blue the results of the analysis at ρ=1, matching 4 of the main widespread substances (ie, heroin, buprenorphine, oxycodone, and fentanyl) with the secondary ROA (upper panel) and the drug-tampering techniques (lower panel). Figure 5 shows the odds ratios of primary ROA and drug-tampering methods. For reference, the green markers represent the ORs obtained at ρ=0 and ρ=∞ for the same categories. Multimedia Appendices 3-5 provide the complete set of results for all the substances identified and the secondary ROA. Due to the low representativeness of intrathecal and urogenital ROA with most of the tampering-related terms, we omitted those categories from the analysis. In the plots, the associations that are not statistically significant (*P*>.01) are reported in gray, and the horizontal lines indicate the OR and the 95% confidence interval. The radius of the circle is proportional to the sample of co-mentions and the dashed vertical line corresponds to an OR of 1 for reference.

Figure 4. Odds ratios of the most widespread opioid substances with routes of administration (top row) and drug-tampering methods (bottom row). Labels on the right axis report the confidence interval at ρ=1. OR: odds ratio.

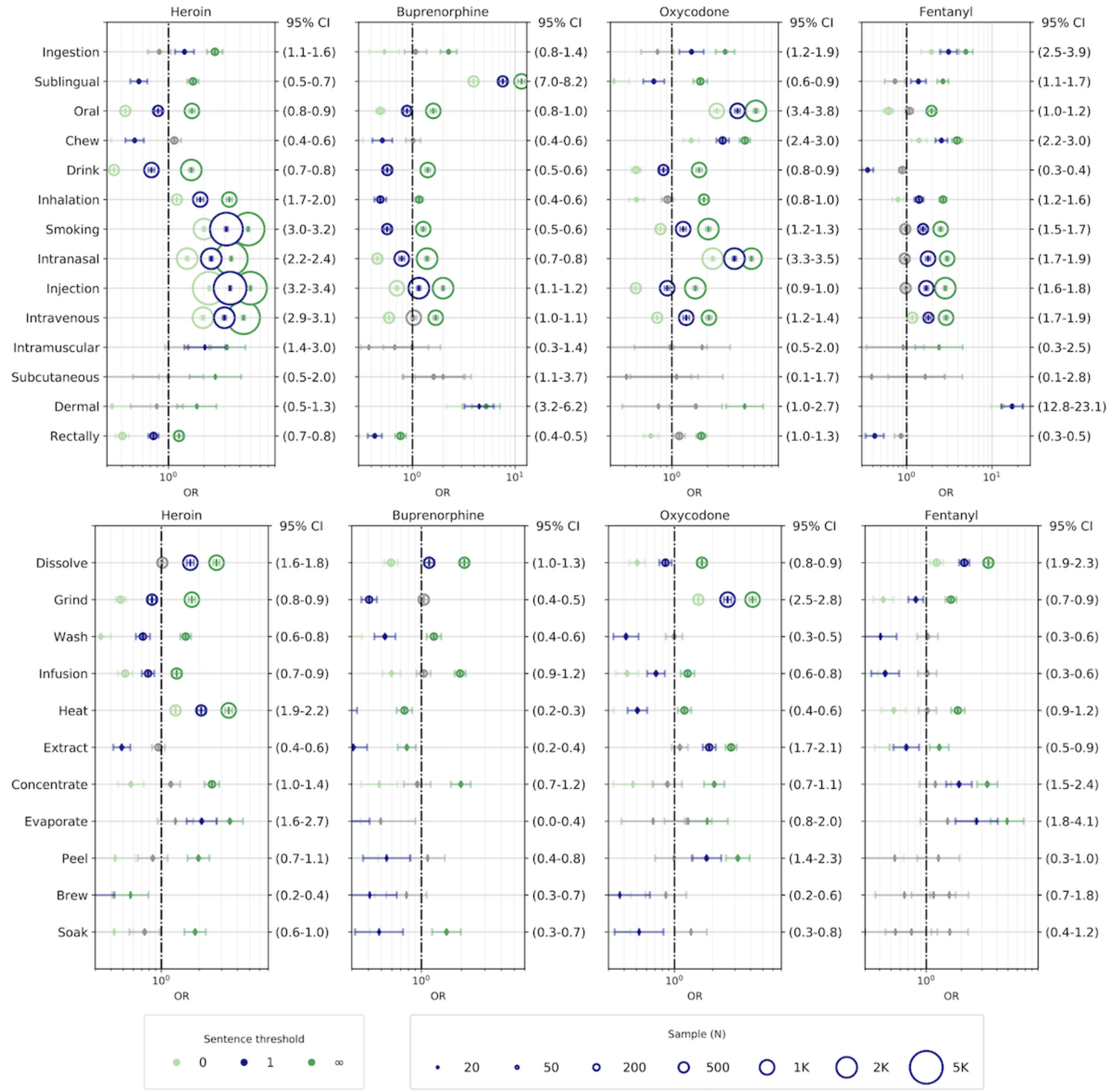
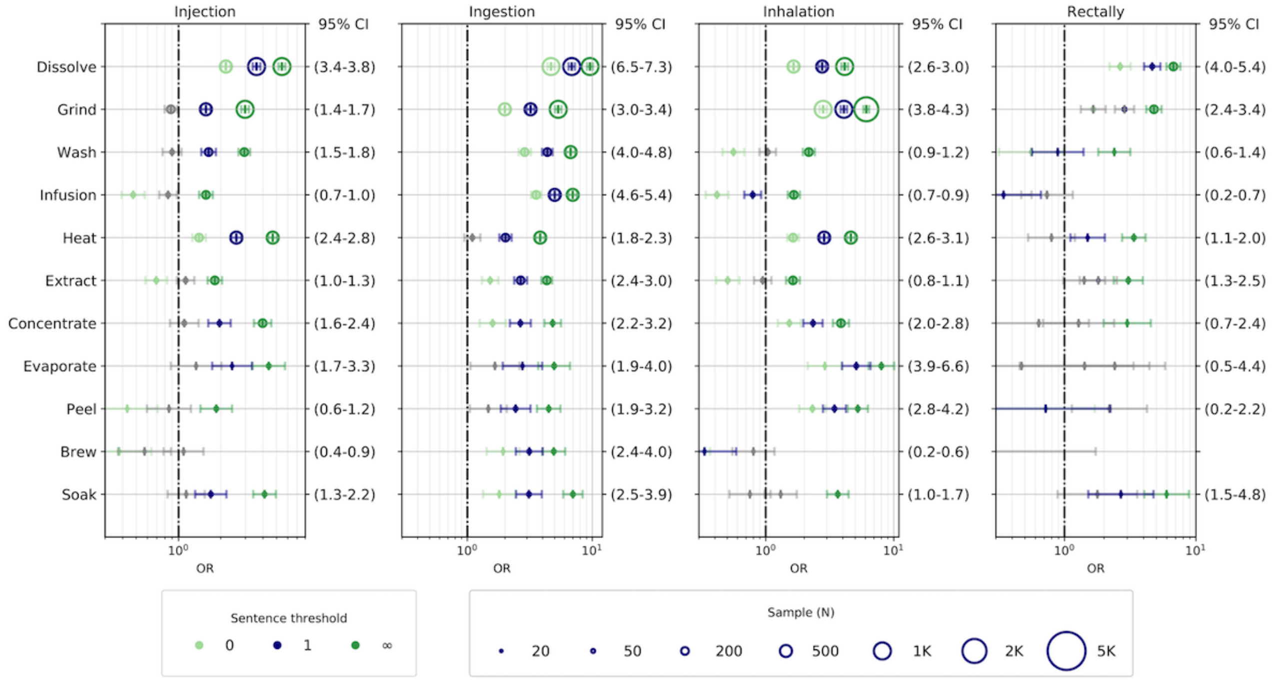


Figure 5. Odds ratios of the primary routes of administration (excluding other routes of administration) and drug-tampering methods. Labels on the right axis report the confidence interval at ρ=1. OR: odds ratio.



Discussion

Opioid Interest on Reddit

In this work, we identified over 3 million comments on 32 subreddits focused on discussing practices and implications of firsthand opioid use. We also selected a cohort of over 86,000 Reddit users interested in this topic. Such a large data set allowed us to assess the magnitude of the online interest in opioids and model its evolution during the 5 years of study, sadly verifying its rapidly increasing popularity. By the end of 2018, the opioid epidemic remained an escalating public health threat, and at the time of writing, the opioid crisis is still calling for countermeasures at scale. Hence, we believe our large data set may constitute a valid alternative source to advise decision making and a valuable starting point for future infodemiology research.

Vocabulary Expansion

By observing the vocabularies in Tables 3-5 resulting from the expansion algorithm, we can ascertain the importance of enriching domain expertise with user-generated content and observe that some common features are captured across categories. Our method was able to detect synonyms and common short names, very specific acronyms (eg, “cwe” for cold water extraction [65]), slang expressions like “sippin” (often used when referring to the act of drinking codeine mixtures [63]), nicknames (eg, “panda” for oxymorphone), and polypharmacy instances (eg, “speedball” and “goofball” [66]). The vocabulary expansion underlines the use of prescription dosages (usually stamped on the tablets) in place of the commercial names of the substances (eg, “30s” for oxycodone). Moreover, we deduced that opioid firsthand users discussed variants of the substances (eg, “bth” and “ecp” for black tar heroin and East Coast powder), research chemical equivalents (eg, “u47700” [67]), and formulations intended for veterinary use (eg, sufentanil, carfentanil).

ROA vocabulary included and categorized both medical terms, adding terms scarcely considered in previous studies, like “vaping,” and nonmedical or unconventional administration terms, such as “chewing,” “snorting,” “smoking,” and “boofing” [39]. Our taxonomy also enabled us to disambiguate common primary ROA, such as injection and ingestion, into specific secondary ones, like subcutaneous [39] and sublingual administrations.

Finally, the drug-tampering vocabulary captured tampering methods that modify the physical status of the substances, like crushing and peeling, and some methods aiming at altering the chemical characteristics of the substances, like dissolving, washing, and heating [41]. We believe that even if this vocabulary might not be exhaustive of all drug-tampering methods, it offers a novel evidence-based perspective on the topic compared with the existing literature. The expanded vocabularies proved essential to fully incorporating the language complexity of online discussions and taboo behaviors [68] into at-scale analyses. Hopefully, our contribution might be useful in the future to find and understand new abusive behaviors that are discussed online, ultimately driving future research to yield more effective prevention methods.

Adoption Popularity of Opioid Substances and ROA

Considering the share of users mentioning a term to be a proxy of firsthand involvement in opioid-related activities and including topic-specific terminology, the longitudinal views in Figures 2 and 3 can be used to rank the popularity of nonmedical usage of opioid substances and ROA and their adoption trends. Ranking the substances by average share, we can see that heroin is by far the most popular substance, mentioned on average by 1 in every 3 users. Its share of users, though, is steadily decreasing, with a loss of 10% reported in state-specific findings by Rosenblum et al [27]. Buprenorphine and oxycodone were the most mentioned prescription opioids; they showed fairly static behavior, while hydrocodone importance decreased over time [28], possibly due to more stringent prescription regulation starting in 2014 [69]. Fentanyl showed the most abrupt behavior, dramatically increasing since 2016. Its volume of mentions in 2018 increased by almost 1.5 times compared with 2014, confirming it as one of the most recent threats [5,28]. In contrast, we did not find evidence of drastic changes in oxymorphone adoption after its partial ban in 2017 [70]. ROA adoption was led by injection and inhalation, which were the most popular ROA across the years, mentioned by 1 of every 3 authors at their peak. These were followed closely by ingestion. Rectal use and other ROA involved, on average, a significantly lower share of users, around 5% and less than 1%, respectively. Nevertheless, rectal administration has shown a sharp increase in popularity since 2016, almost doubling its share. Administration through inhalation was equally staggered by the intranasal and smoking categories of secondary ROA, strong indicators that this route of administration is indeed capturing nonmedical use of opioids.

This work on understanding which substances are currently gaining popularity may give prevention programs a strategic advantage, especially if consumption trends can be localized geographically [12,30,71], focusing the interventions needed to prevent early adoption of emerging dangerous substances like fentanyl. Moreover, tracking the evolution of interest in prescription opioids might be useful for evaluating the efficacy of ban policies, as in the case of oxymorphone. Understanding which ROA are the most adopted might eventually help address targeted campaigns informing users on safer practices, develop better tamper-resistant prescription drugs, and ultimately better inform the health system of the health risks specific to opioid adoption.

Characterizing the Association Between Substance Consumption, ROA, and Drug-Tampering Methods

By jointly considering the results of the odds ratios in Figures 4 and 5 and Multimedia Appendices 5-7, we can outline complex preferences for the nonmedical use of opioids, triangulating substance use, ROA, and drug-tampering methods. We noticed that the majority of substances exhibited more than one high odds ratio, both with ROA and drug-tampering methods, meaning that such substances might be consumed by users in multiple nonexclusive ways. Our analysis shows that for the most part, the expected medical and nonmedical routes of administration of each substance (ie, intended ROA or known abusive administration) had high odds ratios. For prescription opioids, oral (medical) use was often confirmed (eg, oxycodone: OR 3.6, 95% CI 3.4-3.8), while intranasal administration was often the preferred nonmedical ROA, followed by injection, especially intravenous administration (eg, hydromorphone: OR 9.1, 95% CI 8.6-9.8) [32,72]. As expected, heroin appeared to be most likely consumed through injection (OR 3.3, 95% CI 3.2-3.4) or smoking, if heated up on aluminum foil (OR 3.1, 95% CI 3.0-3.2). Heroin was the only substance that showed high correlations with this administration route. It was also reported to be snorted [64].

Besides confirming and quantifying some known behaviors, our analysis can provide additional insights on the nonmedical use of intended routes of administration. In accordance with the literature [31,32,40,73], we found evidence that abuse of prescription opioids may be associated with chewing the pills (eg, oxycodone: OR 2.7, 95% CI 2.4-3.0). From the analysis of ROA and drug-tampering methods, it appears that nonmedical oral administration was correlated with dissolving (OR 9.7, 95% CI 9.0-10.4), grinding, and washing the substances. In some cases, oral and chewing-related misuse of prescription opioids simply consisted of peeling (OR 5.1, 95% CI 2.6-9.9) the external coating, which is usually hard to chew or responsible for the extended-release effect. Even though some formulations, such as Opana ER (oxymorphone hydrochloride extended-release tablets; Endo Pharmaceuticals), are known to be tamper resistant to crushing, users can peel the tablets to get rid of the extended release coating for higher recreational effects. Injection usually requires that the substance be dissolved (OR 3.5, 95% CI 3.2-3.7), while inhalation requires that the substance be ground to powder, especially for intranasal abuse (OR 6.7, 95% CI 6.3-7.1).

Our method ultimately found evidence of unconventional nonmedical administration for most of the substances. We found a high correlation between dissolving and intranasal administration (OR 4.1, 95% CI 3.8-4.4), which may indicate the adoption of “monkey water,” the practice of dissolving soluble substances, like tar heroin and fentanyl patches, and using the resulting liquid as a nasal spray [36]. Fentanyl patches were also consumed in other unforeseen ways; an unexpectedly high OR of fentanyl and chewing (OR 2.6, 95% CI 2.2-3.0) suggests that prescription patches intended for transdermal use may be chewed for nonmedical use. Our analyses revealed the high odds of abuse of codeine via drinking (OR 4.0, 95% CI 3.7-4.3) codeine syrup, made by extracting or brewing the cough suppressants (OR 14.1, 95% CI 11.5-17.2) and forming the so-called lean or purple drank [7,63,74].

Buprenorphine, usually administered sublingually in its formulations without an antagonist, such as Subutex (buprenorphine; Indivior), and orally in combination with naloxone in the form of pills, such as Suboxone (buprenorphine-naloxone; Indivior) and Zubsolv (buprenorphine-naloxone; Orexo), measured exceptionally high odds of sublingual administration (OR 7.6, 95% CI 7.0-8.2). Evidence of nonmedical use of buprenorphine was also found in the association between dissolving and sublingual use (OR 18.9, 95% CI 16.8-21.3). Opioid firsthand users know that the opioid antagonist in buprenorphine-naloxone compounds has low bioavailability if dissolved under the tongue; hence, to achieve higher opioid effects and eliminate the antagonist, these compounds are generally taken sublingually and not through other ROA, with which buprenorphine shows negative associations.

Finally, our study shows that rectal administration is a viable and unforeseen option for the nonmedical use of some opioids, resulting in higher recreational effects, especially with hydromorphone (OR 5.2, 95% CI 4.6-6.0), morphine, and oxymorphone. Rectal administration showed high correlations with the dissolving, grinding, and soaking drug-tampering methods, possible indicators of an unconventional route of administration, largely overlooked, which involves dissolving the substances in liquid water or alcohol (ie, "butt-chugging") [39,75]. Subcutaneous administration was only weakly associated with morphine, suggesting that the practice of "skin popping" [38], which consists of injecting the substance in the tissues under the skin, is potentially not widespread.

The complex interactions between substance use, routes of administration, and drug tampering that can be unveiled with our methodology provide a broad yet detailed perspective on the nonmedical use of opioids, evidencing abusive behaviors in which unconventional ROA and drug tampering play a key role. Knowledge about abusive behaviors could be taken into consideration by physicians during treatment programs, allowing them to favor opioid medications that are less likely to be transformed and abused. Our results should be addressed with effective health policies, driving future clinical research to better focus its efforts on understanding health-related risks and guiding the production of new tamper-resistant and abuse-deterrent opioid formulations.

Limitations and Future Work

We acknowledge some limitations in the present research. The population sampled on Reddit might have intrinsic social media biases, and it is likely not representative of the general population (eg, for gender, age, or ethnicity). Moreover, since we enrolled the users in our cohort based on their engagement in subcommunities focusing on firsthand use of opioids, we cannot exclude the possibility that in some cases, such users might have been reporting secondhand experiences, disseminating general news, or discussing intended medical drug use for pain management. We must also consider that the selected individuals were not clinically diagnosed with opioid use disorder. Future work will be devoted to building a classifier at the user level to identify individuals with opioid use disorder. We are aware that Reddit data have some gaps [76], but since the incompleteness mostly affects the years before 2010, we consider the overall results of our work to not be significantly biased. Other limitations are related to the analytic pipeline, where we narrowed our text analysis to term counts and co-occurrences, which might have produced spillover effects in comments discussing multiple topics and could have amplified the strength of cross-associations. Future work will include n-grams and more context-based language models. Finally, it is worth stressing that the measure of association through odds ratios should not be intended by any means as an indication of causal effects. This work is an observational study focusing on the characterization of a complex and faceted social phenomenon rather than the identification of determinants or interventions, and it shares the strengths and limitations of correlational studies, especially in medical research.

Ethics and Privacy

Given the sensitive nature of the information shared, including users' vulnerabilities and personal information, privacy and ethical considerations are paramount. In this work, we followed the guidelines and directives in Eysenbach and Till [77], which describe recommendations to ethically conduct medical research with user-generated online data, and we relied on the vast experience of research works dealing with sensitive data gathered on social media [47,78-81]. The researchers had no interactions with the users and have no interest in harming any, and the analyses were performed and reported in the spirit of knowledge, prevention, and harm reduction. In this direction, it is worth noting that the subreddits under study are of public domain, are not password protected, and have thousands of active subscribers; users were fully aware of the public nature of the content they posted and of its free accessibility on the web. Moreover, Reddit offers pseudonymous accounts and strong privacy protection, making it it unlikely that the true identity of a user can be recovered. Nevertheless, in order to further protect the privacy and anonymity of the users in our data set, all information about the names of the authors was anonymized before using the data for analysis. Moreover, every analysis performed was intended to provide aggregated estimates aimed at research purposes, and this work did not include any quotes or information that focused on single authors. Following the directives in Eysenbach and Till [77], our research did not require informed consent.

Conclusions

In this work, we characterized opioid-related discussions on Reddit over 5 years, involving more than 86,000 unique users, and focused on firsthand experiences and nonmedical use. To address the complexity of the language in social media communications, especially in the presence of taboo behaviors such as drug abuse, we gathered a large set of colloquial and nonmedical terms that covered most opioid substances, routes of administration, and drug-tampering methods. We were able to characterize the temporal evolution of the discourse and identify notable trends, such as the surge in the popularity of fentanyl and the decrease in the relative interest in heroin. Focusing on routes of administration, we extended pharmacological and medical research with an in-depth characterization of how opioids substances are administered, since different practices imply different effects and potential health-related risks. We proposed a 2-layer taxonomy and corresponding vocabulary that enabled us to study both medical and recreational routes of administration. We demonstrated the presence of conventional nonmedical ROA (eg, intranasal administration and intravenous injection) and the spread of less conventional practices (eg, an increasing trend in rectal use). In particular, with reference to nonconventional ROA, we characterized for the first time at scale the phenomenon of drug tampering, which could have an impact on health outcomes, since it alters the pharmacokinetics of medications. The interplay between these dimensions was systematically characterized by quantitatively measuring the odds ratios, providing an insightful picture of the complex phenomenon of opioid consumption as discussed on Reddit.

Acknowledgments

PB acknowledges support from the Intesa Sanpaolo Innovation Center. The funder had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript. RS was partially supported by the project Countering Online Hate Speech Through Effective On-line Monitoring, funded by Compagnia di San Paolo.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Schematic representation of the structure of Reddit. Reddit's most common access point is the front page, where the most relevant content of the moment is collected. The users can post on already-existing subreddits or they can create and moderate new ones on any topic of choice. In a subreddit, users can either create a new thread via a submission or indefinitely expand the conversation tree by commenting on an existing thread. The level of content moderation in a subreddit is solely decided by its moderators.

[PNG File , 71 KB - jmir_v23i1e21212_app1.png]

Multimedia Appendix 2

Subreddits discussing firsthand nonmedical use of opioids. An X marks the presence of a subreddit in a specific year.

[XLSX File (Microsoft Excel File), 7 KB - jmir_v23i1e21212_app2.xlsx]

Multimedia Appendix 3

Relevant training parameters of the word embeddings. All the other parameters are set to default values. Two state-of-the-art word embedding models, namely word2vec, and GloVe, have been trained with all the comments and submissions in our subreddits dataset. After a-posteriori validation by a domain expert in terms of topical coherence, we choose word2vec as the reference word embedding model.

[XLSX File (Microsoft Excel File), 6 KB - jmir_v23i1e21212_app3.xlsx]

Multimedia Appendix 4

Cumulative probability of finding n or fewer terms in a sentence for submissions and comments (left panel). Cumulative probability of having n or fewer sentences in a submission or a comment (right panel). Plots refer to the selected subreddit in 2018.

[PNG File , 128 KB - jmir_v23i1e21212_app4.png]

Multimedia Appendix 5

Odds Ratios of opioid substances and Secondary Routes of Administration. The central line and the bar mark the OR and the 95% confidence interval respectively, while the size of the circle is proportional to the sample of co-mentions. Measures that are not statistically significant (*P* >.01) are reported in gray. Labels on the right axis report the Odds Ratio and the confidence interval.

[PNG File , 694 KB - jmir_v23i1e21212_app5.png]

Multimedia Appendix 6

Odds Ratios of opioid substances and Drug Tampering Methods. The central line and the bar mark the OR and the 95% confidence interval respectively, while the size of the circle is proportional to the sample of co-mentions. Measures that are not statistically significant (*P* >.01) are reported in gray. Labels on the right axis report the Odds Ratio and the confidence interval.

[PNG File , 577 KB - jmir_v23i1e21212_app6.png]

Multimedia Appendix 7

Odds Ratios of Secondary Routes of Administration and Drug Tampering Methods. The central line and the bar mark the OR and the 95% confidence interval respectively, while the size of the circle is proportional to the sample of co-mentions. Measures that are not statistically significant (*P* >.01) are reported in gray. Labels on the right axis report the Odds Ratio and the confidence interval.

[PNG File , 598 KB - jmir_v23i1e21212_app7.png]

References

1. Centers for Disease Control and Prevention. Drug Overdose Deaths, Centers for Disease Control and Prevention website. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/drugoverdose/data/statedeaths.html [accessed 2020-05-27] [WebCite Cache ID https://www.cdc.gov/drugoverdose/data/statedeaths.html]

2. Kolodny A, Courtwright DT, Hwang CS, Kreiner P, Eadie JL, Clark TW, et al. The prescription opioid and heroin crisis: a public health approach to an epidemic of addiction. Annu Rev Public Health 2015 Mar 18;36:559-574. [doi: 10.1146/annurev-publhealth-031914-122957] [Medline: 25581144]
3. Compton WM, Jones CM, Baldwin GT. Relationship between Nonmedical Prescription-Opioid Use and Heroin Use. N Engl J Med 2016 Jan 14;374(2):154-163. [doi: 10.1056/nejmra1508490]
4. Rose M. Are Prescription Opioids Driving the Opioid Crisis? Assumptions vs Facts. Pain Med 2018 Apr 01;19(4):793-807 [FREE Full text] [doi: 10.1093/pm/pnx048] [Medline: 28402482]
5. Ciccarone D. The triple wave epidemic: Supply and demand drivers of the US opioid overdose crisis. Int J Drug Policy 2019 Sep;71:183-188 [FREE Full text] [doi: 10.1016/j.drugpo.2019.01.010] [Medline: 30718120]
6. McCabe SE, Cranford JA, Boyd CJ, Teter CJ. Motives, diversion and routes of administration associated with nonmedical use of prescription opioids. Addict Behav 2007 Mar;32(3):562-575 [FREE Full text] [doi: 10.1016/j.addbeh.2006.05.022] [Medline: 16843611]
7. Agnich LE, Stogner JM, Miller BL, Marcum CD. Purple drank prevalence and characteristics of misusers of codeine cough syrup mixtures. Addict Behav 2013 Sep;38(9):2445-2449. [doi: 10.1016/j.addbeh.2013.03.020] [Medline: 23688907]
8. Katz N, Fernandez K, Chang A, Benoit C, Butler SF. Internet-based Survey of Nonmedical Prescription Opioid Use in the United States. Clin J Pain 2008;24(6):528-535. [doi: 10.1097/ajp.0b013e318167a087]
9. Butler SF, Budman SH, Licari A, Cassidy TA, Lioy K, Dickinson J, et al. National addictions vigilance intervention and prevention program (NAVIPPRO): a real-time, product-specific, public health surveillance system for monitoring prescription drug abuse. Pharmacoepidemiol Drug Saf 2008 Dec;17(12):1142-1154. [doi: 10.1002/pds.1659] [Medline: 18932173]
10. Butler SF, Black RA, Cassidy TA, Dailey TM, Budman SH. Abuse risks and routes of administration of different prescription opioid compounds and formulations. Harm Reduct J 2011 Oct 19;8:29 [FREE Full text] [doi: 10.1186/1477-7517-8-29] [Medline: 22011626]
11. Curtis HJ, Croker R, Walker AJ, Richards GC, Quinlan J, Goldacre B. Opioid prescribing trends and geographical variation in England, 1998-2018: a retrospective database study. Lancet Psychiatry 2019 Feb;6(2):140-150. [doi: 10.1016/S2215-0366(18)30471-1] [Medline: 30580987]
12. Schifanella R, Vedove DD, Salomone A, Bajardi P, Paolotti D. Spatial heterogeneity and socioeconomic determinants of opioid prescribing in England between 2015 and 2018. BMC Med 2020 May 15;18(1):127 [FREE Full text] [doi: 10.1186/s12916-020-01575-0] [Medline: 32410615]
13. Richards GC, Mahtani KR, Muthee TB, DeVito NJ, Koshiaris C, Aronson JK, et al. Factors associated with the prescribing of high-dose opioids in primary care: a systematic review and meta-analysis. BMC Med 2020 Mar 30;18(1):68 [FREE Full text] [doi: 10.1186/s12916-020-01528-7] [Medline: 32223746]
14. van Amsterdam J, van den Brink W. The Misuse of Prescription Opioids: A Threat for Europe? Curr Drug Abuse Rev 2015;8(1):3-14. [doi: 10.2174/187447370801150611184218] [Medline: 26084418]
15. Brownstein JS, Freifeld CC, Madoff LC. Digital Disease Detection — Harnessing the Web for Public Health Surveillance. N Engl J Med 2009 May 21;360(21):2153-2157. [doi: 10.1056/nejmp0900702]
16. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
17. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]
18. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. J Med Internet Res 2017 Oct 31;19(10):e353 [FREE Full text] [doi: 10.2196/jmir.6426] [Medline: 29089287]
19. Hu H, Phan N, Geller J, Iezzi S, Vo H, Dou D, et al. An Ensemble Deep Learning Model for Drug Abuse Detection in Sparse Twitter-Sphere. Stud Health Technol Inform 2019 Aug 21;264:163-167. [doi: 10.3233/SHTI190204] [Medline: 31437906]
20. Prieto JT, Scott K, McEwen D, Podewils LJ, Al-Tayyib A, Robinson J, et al. The Detection of Opioid Misuse and Heroin Use From Paramedic Response Documentation: Machine Learning for Improved Surveillance. J Med Internet Res 2020 Jan 03;22(1):e15645 [FREE Full text] [doi: 10.2196/15645] [Medline: 31899451]
21. Ertugrul A, Lin Y, Taskaya-Temizel T. CASTNet: Community-Attentive Spatio-Temporal Networks for Opioid Overdose Forecasting. arXiv. 2019. URL: https://arxiv.org/abs/1905.04714?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%253A+arxiv%252FQSXk+%2528ExcitingAds%2521+cs+updates+on+arXiv.org%2529 [accessed 2020-12-11]
22. Lu J, Sridhar S, Pandey R, Hasan M, Mohler G. Redditors in Recovery: Text Mining Reddit to Investigate Transitions into Drug Addiction. arXiv. 2019. URL: https://arxiv.org/abs/1903.04081 [accessed 2020-12-11]
23. Yang Z, Nguyen L, Jin F. Predicting Opioid Relapse Using Social Media Data. arXiv. 2018. URL: https://arxiv.org/abs/1811.12169 [accessed 2020-12-11]
24. Chancellor S, Nitzburg G, Hu A, Zampieri F, De CM. Discovering alternative treatments for opioid use recovery using social media. 2019 May 04 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 124A. [doi: 10.1145/3290605.3300354]

25. Phalen P, Ray B, Watson DP, Huynh P, Greene MS. Fentanyl related overdose in Indianapolis: Estimating trends using multilevel Bayesian models. Addict Behav 2018 Nov;86:4-10. [doi: 10.1016/j.addbeh.2018.03.010] [Medline: 29631798]
26. Zhu W, Chernew ME, Sherry TB, Maestas N. Initial Opioid Prescriptions among U.S. Commercially Insured Patients, 2012-2017. N Engl J Med 2019 Mar 14;380(11):1043-1052 [FREE Full text] [doi: 10.1056/NEJMsa1807069] [Medline: 30865798]
27. Rosenblum D, Unick J, Ciccarone D. The Rapidly Changing US Illicit Drug Market and the Potential for an Improved Early Warning System: Evidence from Ohio Drug Crime Labs. Drug Alcohol Depend 2020 Mar 01;208:107779. [doi: 10.1016/j.drugalcdep.2019.107779] [Medline: 31931266]
28. Black JC, Margolin ZR, Olson RA, Dart RC. Online Conversation Monitoring to Understand the Opioid Epidemic: Epidemiological Surveillance Study. JMIR Public Health Surveill 2020 Jun 29;6(2):e17073 [FREE Full text] [doi: 10.2196/17073] [Medline: 32597786]
29. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz J, et al. Social Media Based Analysis of Opioid Epidemic Using Reddit. AMIA Annu Symp Proc 2018;2018:867-876 [FREE Full text] [Medline: 30815129]
30. Balsamo D, Bajardi P, Panisson A. Firsthand Opiates Abuse on Social Media: Monitoring Geospatial Patterns of Interest Through a Digital Cohort. 2019 Presented at: WWW '19: The World Wide Web Conference; May 13-17, 2019; San Francisco, CA. [doi: 10.1145/3308558.3313634]
31. Kirsh K, Peppin J, Coleman J. Characterization of prescription opioid abuse in the United States: focus on route of administration. J Pain Palliat Care Pharmacother 2012 Dec;26(4):348-361. [doi: 10.3109/15360288.2012.734905] [Medline: 23675595]
32. Gasior M, Bond M, Malamut R. Routes of abuse of prescription opioid analgesics: a review and assessment of the potential impact of abuse-deterrent formulations. Postgrad Med 2016 Jan;128(1):85-96. [doi: 10.1080/00325481.2016.1120642] [Medline: 26566680]
33. Strang J, Bearn J, Farrell M, Finch E, Gossop M, Griffiths P, et al. Route of drug use and its implications for drug effect, risk of dependence and health consequences. Drug Alcohol Rev 1998 Jun;17(2):197-211. [doi: 10.1080/09595239800187001] [Medline: 16203485]
34. Young AM, Havens JR, Leukefeld CG. Route of administration for illicit prescription opioids: a comparison of rural and urban drug users. Harm Reduct J 2010 Oct 15;7:24 [FREE Full text] [doi: 10.1186/1477-7517-7-24] [Medline: 20950455]
35. Katz N, Dart RC, Bailey E, Trudeau J, Osgood E, Paillard F. Tampering with prescription opioids: nature and extent of the problem, health consequences, and solutions. Am J Drug Alcohol Abuse 2011 Jul;37(4):205-217. [doi: 10.3109/00952990.2011.569623] [Medline: 21517709]
36. Ciccarone D. Heroin in brown, black and white: structural factors and medical consequences in the US heroin market. Int J Drug Policy 2009 May;20(3):277-282 [FREE Full text] [doi: 10.1016/j.drugpo.2008.08.003] [Medline: 18945606]
37. Carlson RG, Nahhas RW, Martins SS, Daniulaityte R. Predictors of transition to heroin use among initially non-opioid dependent illicit pharmaceutical opioid users: A natural history study. Drug Alcohol Depend 2016 Mar 01;160:127-134 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.12.026] [Medline: 26785634]
38. Coon TP, Miller M, Kaylor D, Jones-Spangle K. Rectal insertion of fentanyl patches: a new route of toxicity. Ann Emerg Med 2005 Nov;46(5):473. [doi: 10.1016/j.annemergmed.2005.06.450] [Medline: 16271683]
39. Rivers Allen J, Bridge W. Strange Routes of Administration for Substances of Abuse. Am J Psychiatry Residents J 2017 Dec;12(12):7-11. [doi: 10.1176/appi.ajp-rj.2017.121203]
40. McCaffrey S, Manser KA, Trudeau KJ, Niebler G, Brown C, Zarycranski D, et al. The natural history of prescription opioid abuse: A pilot study exploring change in routes of administration and motivation for changes. J Opioid Manag 2018;14(6):397-405. [doi: 10.5055/jom.2018.0472] [Medline: 30629276]
41. Mastropietro DJ. Drug Tampering and Abuse Deterrence. J Dev Drugs 2013;03(1):119. [doi: 10.4172/2329-6631.1000119]
42. Manikonda L, Beigi G, Liu H, Kambhampati S. Twitter for sparking a movement, reddit for sharing the moment: #metoo through the lens of social media. arXiv. 2018. URL: https://arxiv.org/abs/1803.08022 [accessed 2020-12-14] [WebCite Cache ID https://arxiv.org/abs/1803.08022]
43. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The Pushshift Reddit Dataset. arXiv. 2020. URL: https://arxiv.org/abs/2001.08435 [accessed 2020-12-14]
44. Medvedev A, Lambiotte R, Delvenne J. The anatomy of Reddit: An overview of academic research. arXiv. 2018. URL: https://arxiv.org/abs/2001.08435 [accessed 2020-12-14]
45. De Choudhury M, De S. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI.
46. Enes KB, Valadares Brum PP, Oliveira Cunha T, Murai F, Couto da Silva AP, Lobo Pappa G. Reddit weight loss communities: do they have what it takes for effective health interventions? 2018 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '18); December 3-6, 2018; Santiago, Chile.
47. Saha K, Kim SC, Reddy MD, Carter AJ, Sharma E, Haimson OL, et al. The Language of LGBTQ+ Minority Stress Experiences on Social Media. Proc ACM Hum Comput Interact 2019 Nov;3(CSCW):89 [FREE Full text] [doi: 10.1145/3361108] [Medline: 32935081]
48. Baumgartner J. Pushshift Reddit. Pushshift. URL: https://files.pushshift.io/reddit/ [accessed 2020-05-27]

49. Baumgartner J. r/datasets - I have every publicly available Reddit comment for research. Reddit. URL: https://www.reddit.com/r/datasets/comments/3bxlg7/ [accessed 2020-05-27]
50. SpaCy. Spacy industrial-strength Natural Language Processing in Python. Spacy. URL: https://spacy.io/ [accessed 2020-05-27]
51. Barabási AL. The origin of bursts and heavy tails in human dynamics. Nature 2005 May 12;435(7039):207-211. [doi: 10.1038/nature03459] [Medline: 15889093]
52. Malmgren RD, Stouffer DB, Campanharo ASLO, Amaral LAN. On universality in human correspondence activity. Science 2009 Sep 25;325(5948):1696-1700 [FREE Full text] [doi: 10.1126/science.1174562] [Medline: 19779200]
53. Muchnik L, Pei S, Parra LC, Reis SDS, Andrade JS, Havlin S, et al. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. Sci Rep 2013;3:1783 [FREE Full text] [doi: 10.1038/srep01783] [Medline: 23648793]
54. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
55. langdetect. Python Software Foundation. URL: https://pypi.org/project/langdetect/ [accessed 2020-12-14] [WebCite Cache ID https://pypi.org/project/langdetect/]
56. pycld2. Python Software Foundation. URL: https://pypi.org/project/pycld2/ [accessed 2020-10-12]
57. pycld3. Python Software Foundation. URL: https://pypi.org/project/pycld3/ [accessed 2020-10-12]
58. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: NIPS'13: 26th International Conference on Neural Information Processing Systems; Dec 5-8, 2013; Lake Tahoe, NV.
59. Urban Dictionary. Urban Dictionary. URL: https://www.urbandictionary.com [accessed 2020-05-27]
60. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. JOSS 2018 Sep;3(29):861. [doi: 10.21105/joss.00861]
61. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 25-29, 2014; Doha, Qatar.
62. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. 2010 Presented at: 9th Python in Science Conference (SciPy 2010); June 28-July 3, 2010; Austin, TX.
63. Hart M, Agnich LE, Stogner J, Miller BL. 'Me and My Drank:' Exploring the Relationship Between Musical Preferences and Purple Drank Experimentation. Am J Crim Just 2013 Jun 1;39(1):172-186. [doi: 10.1007/s12103-013-9213-7]
64. Surratt HL, Kurtz SP, Buttram M, Levi-Minzi MA, Pagano ME, Cicero TJ. Heroin use onset among nonmedical prescription opioid users in the club scene. Drug Alcohol Depend 2017 Oct 01;179:131-138 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.06.034] [Medline: 28772173]
65. Bausch J, Kershman A, Shear J, Lewis L. Tamper resistant lipid-based oral dosage form for opioid agonists. US Patent 8,273,798. 2012. URL: https://patentimages.storage.googleapis.com/f6/a0/b2/087f721ed668c8/US8273798.pdf [accessed 2020-12-11]
66. Ellis MS, Kasper ZA, Cicero TJ. Twin epidemics: The surging rise of methamphetamine use in chronic opioid users. Drug Alcohol Depend 2018 Dec 01;193:14-20. [doi: 10.1016/j.drugalcdep.2018.08.029] [Medline: 30326396]
67. Prekupec MP, Mansky PA, Baumann MH. Misuse of Novel Synthetic Opioids. J Addict Med 2017;11(4):256-265. [doi: 10.1097/adm.0000000000000324]
68. Allan K, Burridge K. Forbidden words: Taboo and the censoring of language. Cambridge, United Kingdom: Cambridge University Press; 2006.
69. 2014 - Final Rule: Rescheduling of Hydrocodone Combination Products From Schedule III to Schedule II. Drug Enforcement Administration, Department of Justice. URL: https://www.deadiversion.usdoj.gov/fed_regs/rules/2014/fr0822.htm [accessed 2020-05-27] [WebCite Cache ID https://www.deadiversion.usdoj.gov/fed_regs/rules/2014/fr0822.htm]
70. Oxymorphone (marketed as Opana ER) Information. US Food and Drug Administration. 2018. URL: https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/oxymorphone-marketed-opana-er-information [accessed 2020-05-27]
71. Basak A, Cadena J, Marathe A, Vullikanti A. Detection of Spatiotemporal Prescription Opioid Hot Spots With Network Scan Statistics: Multistate Analysis. JMIR Public Health Surveill 2019 Jun 17;5(2):e12110 [FREE Full text] [doi: 10.2196/12110] [Medline: 31210142]
72. Omidian A, Mastropietro D, Omidian H. Routes of Opioid Abuse and its Novel Deterrent Formulations. J Develop Drugs 2015;4(5):e1. [doi: 10.4172/2329-6631.1000141]
73. Butler SF, Cassidy TA, Chilcoat H, Black RA, Landau C, Budman SH, et al. Abuse rates and routes of administration of reformulated extended-release oxycodone: initial findings from a sentinel surveillance sample of individuals assessed for substance abuse treatment. J Pain 2013 Apr;14(4):351-358. [doi: 10.1016/j.jpain.2012.08.008] [Medline: 23127293]
74. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of Codeine Misuse on Instagram: Content Analysis. JMIR Public Health Surveill 2018 Mar 20;4(1):e22 [FREE Full text] [doi: 10.2196/publichealth.8144] [Medline: 29559422]
75. El Mazloum R, Snenghi R, Barbieri S, Feltracco P, Omizzolo L, Vettore G, et al. 'Butt-chugging' a new way of alcohol assumption in young people. Eur J Public Health 2015;25:1. [doi: 10.1093/eurpub/ckv170.089]

76. Gaffney D, Matias JN. Caveat emptor, computational social science: Large-scale missing data in a widely-published Reddit corpus. PLoS One 2018;13(7):e0200162 [FREE Full text] [doi: 10.1371/journal.pone.0200162] [Medline: 29979741]
77. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
78. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):708-713 [FREE Full text] [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
79. Chancellor S, Birnbaum ML, Caine ED, Silenzio VMB, De Choudhury M. A taxonomy of ethical tensions in inferring mental health states from social media. 2019 Presented at: 2019 ACM Conference on Fairness, Accountability, and Transparency; Jan 29-31, 2019; Atlanta, GA.
80. Ramírez-Cifuentes D, Freire A, Baeza-Yates R, Puntí J, Medina-Bravo P, Velazquez DA, et al. Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis. J Med Internet Res 2020 Jul 07;22(7):e17758 [FREE Full text] [doi: 10.2196/17758] [Medline: 32673256]
81. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring Online Discussions About Suicide Among Twitter Users With Schizophrenia: Exploratory Study. JMIR Ment Health 2018 Dec 13;5(4):e11483. [doi: 10.2196/11483]

Abbreviations

OR: odds ratio
ROA: routes of administration
UMAP: uniform manifold approximation and projection

Edited by G Eysenbach; submitted 08.06.20; peer-reviewed by A Roundtree, AM Auvinen; comments to author 01.08.20; revised version received 15.10.20; accepted 28.10.20; published 04.01.21.

Please cite as:
Balsamo D, Bajardi P, Salomone A, Schifanella R
Patterns of Routes of Administration and Drug Tampering for Nonmedical Opioid Consumption: Data Mining and Content Analysis of Reddit Discussions
J Med Internet Res 2021;23(1):e21212
URL: https://www.jmir.org/2021/1/e21212
doi:10.2196/21212
PMID:33393910

Original Paper

“Down the Rabbit Hole” of Vaccine Misinformation on YouTube: Network Exposure Study

Lu Tang[1], PhD; Kayo Fujimoto[2], PhD; Muhammad (Tuan) Amith[3], PhD; Rachel Cunningham[4], MPH; Rebecca A Costantini[1], MA; Felicia York[1], MA; Grace Xiong[5]; Julie A Boom[4], MD; Cui Tao[3], PhD

[1]Department of Communication, Texas A&M University, College Station, TX, United States
[2]School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States
[3]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States
[4]Immunization Project, Texas Children’s Hospital, Houston, TX, United States
[5]Department of Neuroscience, University of Texas, Austin, TX, United States

Corresponding Author:
Cui Tao, PhD
School of Biomedical Informatics
The University of Texas Health Science Center at Houston
7000 Fannin Suite 600
Houston, TX, 77030
United States
Phone: 1 713 500 3981
Email: cui.tao@uth.tmc.edu

Abstract

Background: Social media platforms such as YouTube are hotbeds for the spread of misinformation about vaccines.

Objective: The aim of this study was to explore how individuals are exposed to antivaccine misinformation on YouTube based on whether they start their viewing from a keyword-based search or from antivaccine seed videos.

Methods: Four networks of videos based on YouTube recommendations were collected in November 2019. Two search networks were created from provaccine and antivaccine keywords to resemble *goal-oriented browsing*. Two seed networks were constructed from conspiracy and antivaccine expert seed videos to resemble *direct navigation*. Video contents and network structures were analyzed using the network exposure model.

Results: Viewers are more likely to encounter antivaccine videos through direct navigation starting from an antivaccine video than through goal-oriented browsing. In the two seed networks, provaccine videos, antivaccine videos, and videos containing health misinformation were all found to be more likely to lead to more antivaccine videos.

Conclusions: YouTube has boosted the search rankings of provaccine videos to combat the influence of antivaccine information. However, when viewers are directed to antivaccine videos on YouTube from another site, the recommendation algorithm is still likely to expose them to additional antivaccine information.

(J Med Internet Res 2021;23(1):e23262) doi:10.2196/23262

KEYWORDS

vaccine; misinformation; infodemiology; infodemic; YouTube; network analysis

Introduction

Background

The proliferation of social media has allowed the antivaccine movement to become more influential than at any point in history [1]. Earlier studies have demonstrated that social media platforms such as Pinterest and Twitter are filled with antivaccine information [2,3]. The consumption of antivaccine social media content could negatively impact vaccine attitudes and consequently vaccine uptake [4]. Furthermore, social media can indirectly influence the public by setting the agenda of traditional mass media in vaccine-related controversies [5]. In some cases, social media platforms such as Twitter have even been weaponized to promote antivaccine messages through the use of bots and trolls [6].

YouTube is the largest video-sharing platform in the world with more than 1 billion users. However, it is a hotbed for antivaccine information. Researchers examined 172 YouTube videos related to the human papillomavirus vaccine and concluded that only slightly over 30% of the videos were provaccine [7]. A more recent study of influenza and measles-mumps-rubella vaccine videos on YouTube showed that around 20% and 30% of the videos were antivaccine in nature, respectively [8]. In both studies, antivaccine videos received more views and likes than provaccine videos [7,8].

YouTube has come under criticism because its recommendation algorithm keeps viewers watching videos by suggesting similar videos based on their viewing histories. In other words, YouTube creates filter bubbles where viewers are exposed to repetitive, homogenous, and often biased content, which further reinforces biases and misconceptions. Scholarly attention has been paid to YouTube content promulgating politically extreme ideologies [9], but little is known about the spread of harmful health content such as misinformation about vaccines on this platform. In this study, we jump into the “rabbit hole” driven by YouTube’s recommendation algorithm to explore viewers’ exposure to vaccine-related information and misinformation to determine to what extent, if any, YouTube’s search and recommendation algorithms impact the information to which audiences are exposed.

Diffusion of Information on YouTube

Traditionally, scholars have explained how individuals consume media information using the selective exposure paradigm, which predicts that individuals tend to select media content consistent with their existing beliefs and attitudes [10]. In the era of social media, individuals’ exposure to social media content is largely influenced by the recommendations of their friends or the celebrities they follow or subscribe to [11]. However, one’s friends and associates on social media tend to share similar interests and opinions; consequently, relying on the recommendation of one’s contacts is likely to create echo chambers, where one is exposed to conforming opinions repeatedly [11]. In addition, individuals’ consumption of social media content is influenced by the machine learning–based recommendation algorithms of the platform. Such personalization of content creates filter bubbles, in which algorithms recommend information that users have previously been exposed to and with which they agree [11]. Empirical studies have shown that both echo chambers and filter bubbles deepen the ideological divide among the public [11].

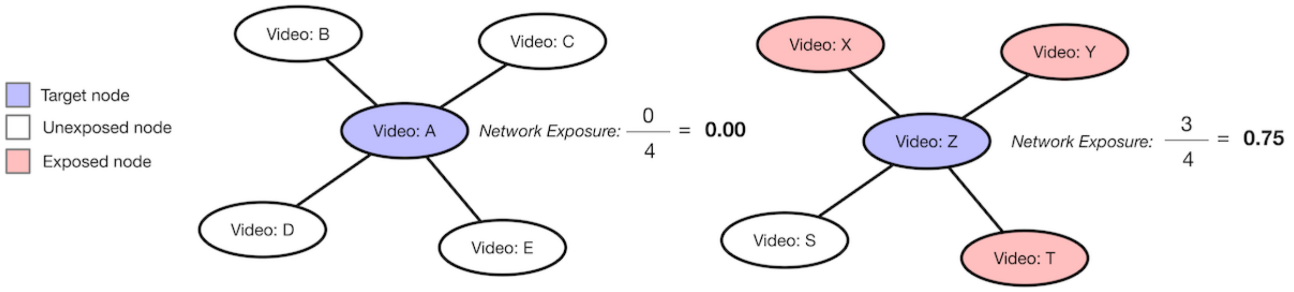
Information Exposure on YouTube

Users interact with the YouTube platform in two ways. First, *direct navigation* occurs when users are directed to watch a YouTube video from another website or social media platform. Alternatively, users could search for videos based on keywords (*goal-oriented browsing*) [12]. In both cases, YouTube will present users with a set of recommended videos based on the user’s prior viewing behaviors as well as covisitation counts [13]. To understand the type of vaccine information YouTube users are exposed to when they use direct navigation or goal-oriented browsing, we propose the first research question (RQ1): When YouTube users start their viewing with provaccine or antivaccine keywords, or an antivaccine seed video, to what extent will they will be exposed to pro- and antivaccine content?

Thus, RQ1 explores YouTube users’ exposure to pro- and antivaccine videos on a macro level when they use YouTube in different ways; however, it is also important to examine the connections among different types of videos in recommendation networks on a micro level. Based on the diffusion of innovation theory [14], the network exposure model (NEM) measures the degree to which a node in the network is exposed to other nodes with a certain attribute. A node’s exposure to an attribute is computed from the average edge from a node to other nodes that exhibit the attribute. In Figure 1 (adapted from Valente [15]), we present two examples with nodes that have connections to other nodes. Some of the nodes (representing any type of entity such as a YouTube video) in Figure 1 are color-coded in pink to indicate that it has an attribute (such as containing misinformation). In Figure 1, Video A has an exposure value of 0 because it has no connections to nodes containing misinformation (ie, 0/4=0.00). In comparison, Video Z has a network exposure value of 0.75, because three out of four edges connect it to nodes (pink nodes) containing misinformation (ie, 3/4=0.75). A more detailed description of the metric can be found in the related literature [15-18].

Overall, the NEM measures the extent to which one node in the network is exposed to a certain type of node. To understand how likely provaccine and antivaccine videos, as well as other types of videos unrelated to vaccines, are to lead to antivaccine information, we addressed RQ2: What is the degree of exposure of pro- and antivaccine videos as well as other videos unrelated to vaccines to additional antivaccine videos?

Figure 1. Network exposure model example showing two nodes (A and Z) exposed to attributes based on their ties. The example was adapted from Valente [15].



Methods

Data Collection and Network Generation

To collect YouTube videos based on a keyword search (*goal-oriented browsing*), we used four provaccine key phrases (“why I vaccinate,” “vaccinate with me,” “vaccine saves lives,” and “vaccine works”), and four antivaccine key phrases (“vaccine causes autism,” “vaccine kills,” “vaccine takes life,” and “vaccine harm”). These key phrases were based on the most popular positive and negative hashtags about vaccines on Twitter in October 2019. We utilized CAS^{2}T [19], an open-sourced tool that leverages the YouTube application programming interface (API) [20] to create networks of related YouTube videos and retrieve each video’s recommended videos. The API does not factor in the viewing histories of individual users and only retrieves videos based on YouTube recommendation algorithms. This tool also stores the collections of videos and their recommended videos in a relational database (SQLite), along with their metadata information (ie, views, likes). For the two networks based on provaccine and antivaccine key phrases (“search networks”), we collected the first 6 videos’ URLs and fetched three depth levels of related videos. Through this procedure, we gathered 6 related videos for each of the videos, and then the related videos’ related videos, and so on. We presumed that users might not view more than 6 recommended videos due to the screen size of typical computers and the number of recommended videos shown on one screen. We performed the same procedure with the videos of referral sharing (“seed network”) for vaccine-based conspiracy videos (ie, “big pharma money-making scheme,” “government covering up side-effects,” “distrust of doctors”) and antivaccine experts claiming authority on the topic of vaccines. The *conspiracy seed network* was seeded with 16 videos and the *antivaccine expert seed network* was seeded with 8 videos. Seed videos were sourced from two playlists on their respective topics. Thus, this provided a good start to further assess the misinformation rabbit hole through referral sharing. Data collection was conducted in November 2019. After data collection, we aggregated the databases from CAS^{2}T into one central database repository using PostgreSQL v12, and assigned a unique identification number to each video.

Annotation

To annotate the two search networks, we started with 815 videos that were initially downloaded. After deleting replicated videos, a total of 538 videos were annotated. We annotated these videos in terms of (1) whether the video was related to vaccines and (2) if the video was related to vaccines, whether it was provaccine or antivaccine. For videos that were unrelated to vaccines, we annotated (1) whether it was about autism and (2) whether it contained other health information and misinformation. Three of the authors (LT, RAC, FY) annotated 54 videos (10%) selected through random systematic sampling and achieved excellent intercoder reliability according to Krippendorff α values (related to vaccine, α=.949; pro- or antivaccine, α=.90; containing autism-related information, α=.96; and containing other health misinformation, α=.949). The three authors then split all of the videos and annotated them independently. Videos in a language other than English were identified as such and excluded from data analysis. We annotated the two seed networks (conspiracy and antivaccine expert networks) using the guidelines described at the beginning of this subsection. Two of the authors (TA, GX) annotated the 1034 seed-based videos, demonstrating excellent reliability (vaccine-related, α=.899; health-based, α=.901; autism-related, α=.96; and misinformation, α=.806). Videos related to vaccines and that contained misinformation were then recoded as “antivaccine videos” and other vaccine-related videos were recoded as “provaccine videos” to create consistency between the coding results of the two groups.

Data Analysis

We analyzed the four networks using various network metrics to understand the relationships among different types of videos within their respective networks. CAS^{2}T conveniently generates tables for node metadata, nodes, and edges, which, for the nodes and edges, can be seamlessly imported into Gephi to construct undirected networks to perform network analysis. Gephi [21] was used to compute the network statistics and generate the visualizations of the networks.

For each of the four networks, we examined how likely nonvaccine videos, vaccine-related videos (provaccine and other antivaccine videos), autism videos, and health-related videos (unrelated to vaccines) are likely to be exposed to antivaccine videos. We used NET-EXPO [22], a Gephi plugin, to compute network exposure. STATA v15 and SPSS v26 were used for basic statistical calculations and frequency computations. We calculated the case-control odds ratio to measure how likely different types of videos (provaccine videos, antivaccine videos, autism videos unrelated to vaccines, health videos unrelated to vaccines, and health misinformation unrelated to vaccines) were exposed to antivaccine videos.

Results

Four network datasets were generated based on provaccine and antivaccine search key phrases (“search network”) and on conspiracy and antivaccine expert seed videos (“seed network”). Each node in these networks represents a video and each edge represents a recommendation relationship (see Table 1 for the descriptive statistics of the four networks). The node size ranged from 283 to 551, and the number of edges ranged from 342 to 671. The average degree for all of the networks was approximately 2.4, except for the conspiracy network, which was 2.3. The average clustering coefficient, which measures the clustering of the nodes in a network, was 0.415 and 0.411 for provaccine and antivaccine search networks respectively, and was 0.06 and 0.1 for the conspiracy seed network and antivaccine expert seed network, respectively.

Table 1. Global statistics for the four collected networks.

Characteristic	Search networks		Seed networks	
	Provaccine search network	Antivaccine search network	Conspiracy seed network	Antivaccine expert seed network
Network characteristics				
Nodes, n	283	354	483	551
Edges, n	342	417	551	671
Average degree	2.4	2.4	2.3	2.4
Network diameter	8	8	14	12
Average clustering coefficient	0.415	0.411	0.06	0.1
Video type, n (%)				
Nonvaccine-related	242 (86)	315 (89)	413 (86)	511 (93)
Vaccine-related	41 (14)	40 (11)	70 (14)	40 (7)
Provaccine[a]	38 (93)	35 (87.5)	34 (49)	15 (38)
Antivaccine[a]	3 (7)	5 (12.5)	36 (51)	25 (63)
Source of videos, n (%)[a]				
Government agencies	23 (56)	14 (35)	0 (0)	0 (0)
Academic institutions and hospitals	6 (15)	13 (33)	9 (13)	1 (3)
Pharmaceutical companies and for-profit organizations	1 (2)	0 (0)	1 (1)	0 (0)
Consumer-generated	3 (7)	5 (13)	33 (47)	26 (65)
News media	8 (20)	9 (23)	27 (39)	13 (33)
Professional associations	0 (0)	2 (5)	0 (0)	0 (0)
Other	0 (0)	2 (5)	0 (0)	0 (0)
Autism-related video, n (%)	6 (2)	21 (6)	13 (3)	22 (4)
Health-related video, n (%)	100 (35)	142 (40)	267 (55)	316 (57)
Accurate health information, n (%)[b]	99 (99)	139 (98)	157 (59)	172 (54)
Health misinformation, n (%)[b]	1 (1)	3 (2)	110 (41)	144 (46)

[a] Percentages are based on the number of vaccine-related videos in a given network.

[b] Percentages are based on the number of health-related videos in a given network.

RQ1 asked whether starting with provaccine and antivaccine keywords, and starting from antivaccine seed videos will lead viewers to pro- or antivaccine information. In the two search networks generated from provaccine and antivaccine keywords, an overwhelming majority of the vaccine-related videos were provaccine (Table 1). Most of these vaccine-related videos were created and uploaded by credible sources such as governmental agencies, hospitals, and academic institutions. In contrast, viewers were much more likely to be exposed to antivaccine information in the conspiracy seed network and in the antivaccine expert seed network. Approximately half of the vaccine-related videos in the two seed networks were consumer-generated (Table 1).

RQ2 examined the degree of exposure of different types of videos to additional antivaccine information. Table 2 presents the odds ratios for exposure to antivaccine misinformation and Figure 2 provides visualizations of the four networks.

In the search networks, nonvaccine videos had a low likelihood of being exposed to antivaccine videos within their respective networks (provaccine and antivaccine search networks). However, vaccine-related videos had a higher chance of being exposed to antivaccine videos. Within the vaccine-related videos of the search networks, provaccine video nodes were more likely to be exposed to antivaccine videos in both networks. However, none of these odds ratios was statistically significant. The only significant results in these two networks were found for the antivaccine search network, autism videos, and health-related videos (whether contentious or noncontentious), which had no chance of being exposed to antivaccine videos. However, the picture was very different in the two seed networks.

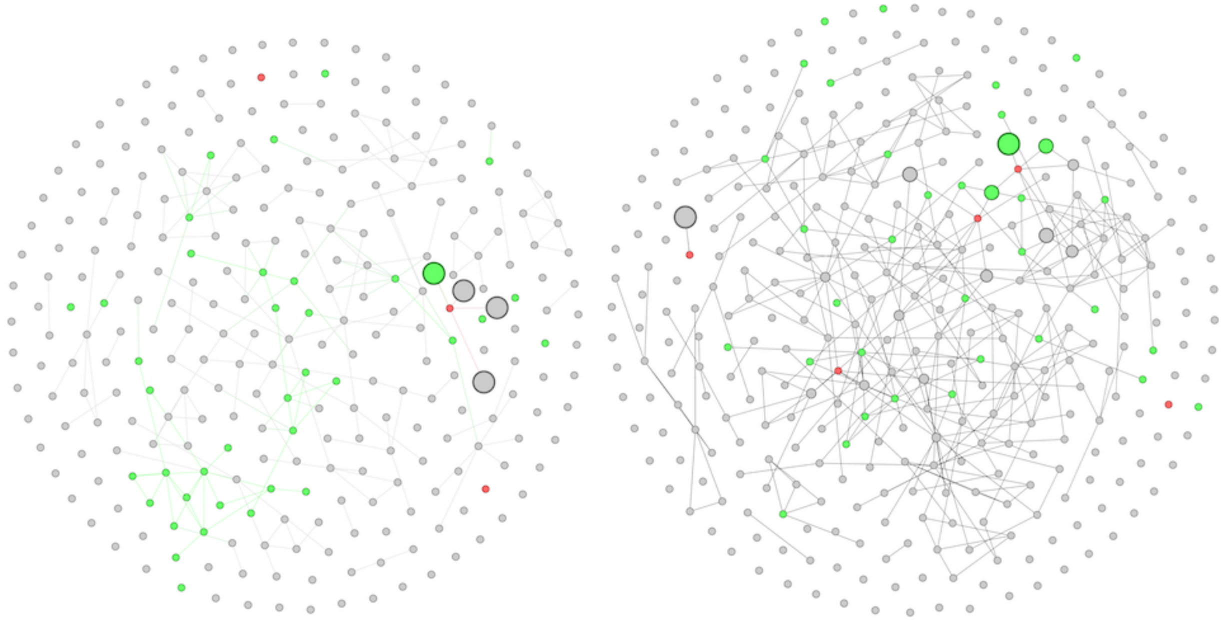
In seed networks, compared to vaccine-related videos, videos unrelated to vaccines had a significantly lower chance of being exposed to antivaccine videos (for both the conspiracy seed network and antivaccine expert seed networks). Antivaccine videos had tremendous exposure to other antivaccine videos (in both the conspiracy seed network and antivaccine expert seed network). In comparison, provaccine videos were also vulnerable to antivaccine video recommendations, but to a lesser extent. Furthermore, videos containing nonvaccine health misinformation had greater odds of being exposed to antivaccine videos in both the conspiracy seed network and the antivaccine expert seed network. In comparison, videos containing accurate health information did not exhibit this pattern.

Table 2. Descriptive statistics of the exposure to antivaccine video nodes.

Statistic	Search networks		Seed networks	
	Provaccine search network	Antivaccine search network	Conspiracy seed network	Antivaccine expert seed network
Mean (SD)	0.01 (0.12)	0.02 (0.10)	0.12 (0.28)	0.07 (0.21)
Range	1-1	0.11-1	0.13-1	0.13-1
Nodes exposed, n (%)	4 (1.4)	15 (4.2)	119 (24.7)	86 (15.6)
Nodes unexposed, n (%)	279 (98.6)	339 (95.8)	364 (75.3)	465 (84.4)
Odds ratio (95% CI)				
Nonvaccine video	0.50 (0.04-27.0)	0.48 (0.12-2.8)	0.07 (0.04-0.14)	0.04 (0.02-0.09)
Vaccine video	1.99 (0.04-25.0)	2.1 (0.36-8.3)	13.6 (7.3-25.9)	24.4 (10.8-58.4)
Provaccine video	2.18 (0.04-27.9)	2.4 (0.41-9.5)	8.94 (3.9-21.6)	12.1 (3.6-46.1)
Antivaccine video	0.00 (0-108.7[a])	0.00 (0.0-18.1[a])	11.6 (5.0-28.8)	27.9 (9.6-97.3)
Autism video	0.00 (0-50.3[a])	0.00 (0.0-4.0[a])	0.92 (0.16-3.6)	2.1 (0.65-5.9)
Health video	5.62 (0.44-297)	0.00 (0.0-0.36[a])	1.52 (0.98-2.4)	2.0 (1.2-3.5)
Accurate health information	5.71 (0.45-301.8)	0.00 (0.0-0.37[a])	0.97 (0.60-1.5)	1.22 (0.72-2.0)
Health misinformation	0.00 (0.00-0.00)	0.00 (0.0-30.6[a])	1.80 (1.1-2.9)	1.76 (1.0-2.9)

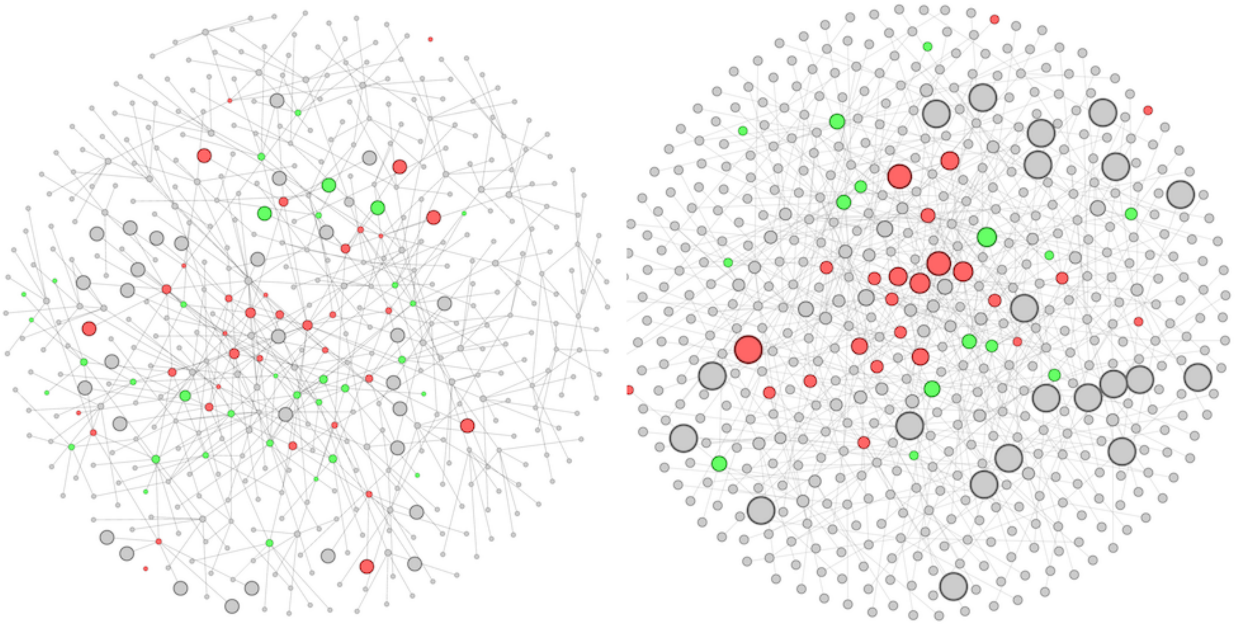
[a] Cornfield exact CI.

Figure 2. Visualization of the four YouTube video networks, where each node represents a video and an edge represents a related link. Green represents a provaccine video and red represents an antivaccine video. Gray nodes are nonvaccine-related videos. The size of the nodes depicts the exposure value to antivaccination videos. Visualization was generated with Gephi.



Pro-vaccine search network

Anti-vaccine search network



Vaccine conspiracy seed network

Anti-vaccine expert seed network

Discussion

Principal Findings

This study tested the characteristics of YouTube’s search and recommendation algorithms by exploring the information users are likely to be exposed to when they begin with keyword-based searches and when they start from an antivaccine seed video. We utilized network exposure models, along with other statistical methods, to determine how vulnerable the videos (and by proxy, the users) are to antivaccine content. Four networks—provaccine search network, antivaccine search network, conspiracy seed network, and antivaccine expert seed network—were examined in terms of both video content and network structures.

First, when users start with a keyword-based search on YouTube, they are likely to reach provaccine videos posted by credible sources such as government agencies and hospitals, regardless of whether they used provaccine or antivaccine keywords. This encouraging finding suggests that YouTube has taken some measures to promote provaccine videos from credible sources in their search function. This contradicts prior research that reported that antivaccine videos appear in the top search hits [23]. Antivaccine users such as Alex Jones, InfoWars, and the like have been banned from YouTube, taking with them their controversial and extreme content [24]. Although certain major players of the antivaccine movement have been removed, there were still several questionable “experts” that emerged in our network data. These antivaccine experts use misleading rhetoric about vaccines that may sound plausible to naïve viewers. Even when such antivaccine experts' videos might not be easily found through keyword searches, they can be found and shared with a network of friends through email or social media, directing the viewers to the proverbial "rabbit hole" of misinformation.

Second, even if users are to watch a provaccine video, they have a relatively high chance of being recommended an antivaccine video. A viewer could have a roughly 2 to 12 times chance of being recommended an antivaccine video depending on the networks than any other video. Antivaccine videos are much more likely to lead to more antivaccine videos. It should be noted that such patterns were statistically significant in the two seed networks but not in the two search networks, which is likely to be caused by the extremely low percentages of antivaccine videos in the two search networks. Furthermore, compared with vaccine-related videos, nonvaccine videos were more isolated from antivaccine content in all four networks. It is assumed that YouTube uses a machine-learning algorithm that looks at titles, descriptions, and other metadata, and matches them with similar data in other videos to denote recommended videos. Therefore, it may not be surprising that nonvaccine videos are somewhat isolated from vaccine videos and that antivaccine videos beget more antivaccine videos.

Third, we found that 2%-6% of the videos in these networks were about autism but did not contain any vaccine-related information. This means that when users watch autism-related videos, they might be directed to vaccine-related videos or even to antivaccine videos. This could potentially contribute to the spread of misinformation about the vaccine-autism link. Furthermore, health videos that discuss nonvaccine health topics (eg, diet, holistic medicine, cancer), especially those that contain misinformation, have some vulnerability to antivaccine videos through recommendations. In the conspiracy and antivaccine expert seed networks, a viewer may come across an antivaccine video while watching a health-related video containing other types of misinformation.

Limitations and Future Directions

One limitation of the current study is the use of the YouTube API. Due to its proprietary nature, it is difficult to ascertain how and why YouTube recommends videos. It is assumed that such recommendations are based on a combination of machine-learning approaches that factor in video metadata, the similarity of words in titles and descriptions with other videos, user preferences, and viewing history. Our networks were based on YouTube recommendations without personal viewing history. In reality, recommended videos may vary with viewers' history and preferences that are included in the calculation. At best, the current study provides a baseline that does not factor in users' viewing history or other preferences. In addition, we came across a considerable number of videos with health information unrelated to vaccines in the networks. It appears that misinformation will lead to additional misinformation. Thus, future research could investigate the content that propagates fad diets, questionable food supplements, and other unsubstantiated health information on YouTube.

Conclusions

This study explored the information YouTube users are exposed to when they start with search keywords or start with seed videos through social network analysis. Utilizing the NEM, we examined the odds of vaccine misinformation propagated to users through individual nodes. Our results showed that although vaccine misinformation is sandboxed within a network of vaccine videos, some provaccine videos are susceptible to antivaccine videos through YouTube recommendations. There is also evidence that health-related videos, especially those containing health misinformation, are vulnerable to vaccine misinformation. Overall, individuals watching YouTube videos through a goal-oriented search have a lower chance of encountering vaccine misinformation due to efforts from YouTube; however, antivaccine misinformation still exists and users have a chance to encounter these videos or other misinformation content through direct navigation.

Acknowledgments

This research was supported by the National Library of Medicine of the National Institutes of Health (R01LM011829), National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R01AI130460 and R01AI130460-03S1), and the National Institute on Alcohol Abuse and Alcoholism (1K99AA019699).

Authors' Contributions

MA conceptualized the research. LT, MA, and KF designed the methodological approach. LT, RAC, FY, RC, and JB developed the annotation classification. LT, RAC, FY, MA, and GX collected and annotated the data. LT, MA, and KF analyzed the data. LT and MA developed the initial draft. All authors read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Wilson K, Keelan J. Social media and the empowering of opponents of medical technologies: the case of anti-vaccinationism. J Med Internet Res 2013 May 28;15(5):e103 [FREE Full text] [doi: 10.2196/jmir.2409] [Medline: 23715762]
2. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. Vaccine 2015 Sep 22;33(39):5051-5056. [doi: 10.1016/j.vaccine.2015.08.064] [Medline: 26319742]

3. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Soc Sci Med 2017 Oct;191:168-175 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.041] [Medline: 28926775]
4. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
5. Jang SM, Mckeever BW, Mckeever R, Kim JK. From Social Media to Mainstream News: The Information Flow of the Vaccine-Autism Controversy in the US, Canada, and the UK. Health Commun 2019 Jan 13;34(1):110-117. [doi: 10.1080/10410236.2017.1384433] [Medline: 29028371]
6. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
7. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health Commun 2012 Jul;27(5):478-485. [doi: 10.1080/10410236.2011.610258] [Medline: 22029723]
8. Yiannakoulias N, Slavik CE, Chase M. Expressions of pro- and anti-vaccine sentiment on YouTube. Vaccine 2019 Apr 03;37(15):2057-2064. [doi: 10.1016/j.vaccine.2019.03.001] [Medline: 30862365]
9. Lewis R. Alternative influence: Broadcasting the reactionary right on YouTube. New York: Data & Society Research Institute; 2018. URL: https://datasociety.net/output/alternative-influence/ [accessed 2020-12-20]
10. Hart W, Albarracín D, Eagly AH, Brechan I, Lindberg MJ, Merrill L. Feeling validated versus being correct: a meta-analysis of selective exposure to information. Psychol Bull 2009 Jul;135(4):555-588 [FREE Full text] [doi: 10.1037/a0015701] [Medline: 19586162]
11. Flaxman S, Goel S, Rao JM. Filter Bubbles, Echo Chambers, and Online News Consumption. Public Opin Q 2016 Mar 22;80(S1):298-320. [doi: 10.1093/poq/nfw006]
12. Davidson J, Livingston B, Sampath D, Liebald B, Liu J, Nandy P, et al. The YouTube video recommendation system. 2010 Sep 26 Presented at: The fourth ACM Conference on Recommender Systems; 2010; Barcelona.
13. O'Callaghan D, Greene D, Conway M, Carthy J, Cunningham P. Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems. Soc Sci Comput Rev 2014 Oct 16;33(4):459-478. [doi: 10.1177/0894439314555329]
14. Rogers E. Diffusion of innovations. New York: Simon and Schuster; 2010.
15. Valente TW. Network models and methods for studying the diffusion of innovations. In: Scott J, Wasserman S, Carrington PJ, editors. Models and methods in social network analysis. New York: Cambridge University Press; 2005:98-116.
16. Marsden PV, Friedkin NE. Network Studies of Social Influence. Sociol Methods Res 2016 Jun 29;22(1):127-151. [doi: 10.1177/0049124193022001006]
17. Valente TW. Network Models of the Diffusion of Innovations. New Jersey: Hampton Press; 1995.
18. Burt RS. Social Contagion and Innovation: Cohesion versus Structural Equivalence. Am J Sociol 1987 May;92(6):1287-1335. [doi: 10.1086/228667]
19. Amiriparian S, Pugachevskiy S, Cummins N, Hantke S, Pohjalainen J, Keren G, et al. CAST a database: Rapid targeted large-scale big data acquisition via small-world modelling of social media platforms. 2017 Oct 23 Presented at: Seventh International Conference on Affective Computing and Intelligent Interaction (ACII); 2017; San Antonio, Texas.
20. Google L. Youtube API. 2019 Nov 10. URL: https://developers.google.com/youtube/v3 [accessed 2019-12-01]
21. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. 2009 May 17 Presented at: Third international AAAI conference on weblogs and social media; 2009; San Jose, CA. [doi: 10.1007/978-1-4614-6170-8_299]
22. Amith M, Fujimoto, K, Tao C. NET-EXPO: A Gephi Plugin Towards Social Network Analysis of Network Exposure for Unipartite and Bipartite Graphs. 2019 Presented at: HCI International 2019: 21st international conference, 2019; July 26-31, 2019; Orlando, FL p. 3-12. [doi: 10.1007/978-3-030-23525-3_1]
23. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm--an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]
24. The YouTube Team. Our ongoing work to tackle hate. YouTube Official Blog. 2019 Jun 05. URL: https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html [accessed 2020-12-13]

Abbreviations

API: application programming interface
NEM: network exposure model

Edited by G Eysenbach; submitted 06.08.20; peer-reviewed by N Yiannakoulias, X Jing; comments to author 10.09.20; revised version received 16.09.20; accepted 16.09.20; published 05.01.21.

Please cite as:
Tang L, Fujimoto K, Amith M(, Cunningham R, Costantini RA, York F, Xiong G, Boom JA, Tao C
"Down the Rabbit Hole" of Vaccine Misinformation on YouTube: Network Exposure Study
J Med Internet Res 2021;23(1):e23262
URL: https://www.jmir.org/2021/1/e23262
doi:10.2196/23262
PMID:33399543

Original Paper

A Novel Machine Learning Framework for Comparison of Viral COVID-19–Related Sina Weibo and Twitter Posts: Workflow Development and Content Analysis

Shi Chen[1,2], PhD; Lina Zhou[3], PhD; Yunya Song[4], PhD; Qian Xu[5], PhD; Ping Wang[6], MSc; Kanlun Wang[3], MSc; Yaorong Ge[7], PhD; Daniel Janies[8], PhD

[1]Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States
[2]School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, United States
[3]School of Business, University of North Carolina at Charlotte, Charlotte, NC, United States
[4]Department of Journalism, Hong Kong Baptist University, Hong Kong, Hong Kong
[5]School of Communications, Elon University, Elon, NC, United States
[6]Department of Medical Informatics, School of Public Health, Jilin University, Jilin, China
[7]Department of Software and Information System, University of North Carolina at Charlotte, Charlotte, NC, United States
[8]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, United States

Corresponding Author:
Shi Chen, PhD
Department of Public Health Sciences
University of North Carolina at Charlotte
9021 University City Blvd
Charlotte, NC, 28223-0001
United States
Phone: 1 8148800738
Email: schen56@uncc.edu

Abstract

Background: Social media plays a critical role in health communications, especially during global health emergencies such as the current COVID-19 pandemic. However, there is a lack of a universal analytical framework to extract, quantify, and compare content features in public discourse of emerging health issues on different social media platforms across a broad sociocultural spectrum.

Objective: We aimed to develop a novel and universal content feature extraction and analytical framework and contrast how content features differ with sociocultural background in discussions of the emerging COVID-19 global health crisis on major social media platforms.

Methods: We sampled the 1000 most shared viral Twitter and Sina Weibo posts regarding COVID-19, developed a comprehensive coding scheme to identify 77 potential features across six major categories (eg, clinical and epidemiological, countermeasures, politics and policy, responses), quantified feature values (0 or 1, indicating whether or not the content feature is mentioned in the post) in each viral post across social media platforms, and performed subsequent comparative analyses. Machine learning dimension reduction and clustering analysis were then applied to harness the power of social media data and provide more unbiased characterization of web-based health communications.

Results: There were substantially different distributions, prevalence, and associations of content features in public discourse about the COVID-19 pandemic on the two social media platforms. Weibo users were more likely to focus on the disease itself and health aspects, while Twitter users engaged more about policy, politics, and other societal issues.

Conclusions: We extracted a rich set of content features from social media data to accurately characterize public discourse related to COVID-19 in different sociocultural backgrounds. In addition, this universal framework can be adopted to analyze social media discussions of other emerging health issues beyond the COVID-19 pandemic.

(J Med Internet Res 2021;23(1):e24889) doi:10.2196/24889

KEYWORDS

COVID-19; Twitter; Sina Weibo; content feature extraction; cross-cultural comparison; machine learning; social media; infodemiology; infoveillance; content analysis; workflow; communication; framework

Introduction

Social media platforms are important communication channels for public engagement of various health issues [1-4]. Through social media, the public can not only receive information from health agencies and news outlets about various health issues [5] but also actively participate in web-based discussions with peers and influencers to exchange opinions about these issues [6]. Social media platforms have been adopted in various health campaigns by both health agencies and concerned groups, including promotion of vaccination [7,8], exercise and healthy lifestyles, and smoking cessation [9].

During health emergencies, especially global infectious disease pandemics, social media has been used substantially by both individuals and organizations. Social media platforms were frequently used during previous public health emergencies of international concern (PHEICs), such as the 2014 Ebola outbreak [10,11] and the 2016 Zika pandemic [12,13]. Social media has also been intensively used during the current COVID-19 pandemic; *COVID-19* is currently the most mentioned keyword across all major social media platforms worldwide. Therefore, social media can be used for infodemiology studies [14-17] to better understand public concerns and make informed decisions regarding the COVID-19 pandemic as well.

Health emergencies are seldom an isolated health or medical issue. Pandemics, including the current COVID-19 pandemic, are almost always intermingled with complicated interactions of underlying societal and cultural factors that vary within and among countries. Consequently, discussions of these pandemics on social media include content not restricted to health, as observed during the 2014 Ebola and 2016 Zika epidemics [10,18-25]. During the current COVID-19 pandemic, it has also been demonstrated that various social and political issues are associated with the pandemic, including different views on nonpharmaceutical interventions (NPIs) such as mask-wearing, social distancing, and stay-at-home-orders [26-29].

To extract and analyze various content features in social media posts, natural language processing (NLP) methods such as linguistic inquiry and word count (LIWC) are usually applied [30]. However, although LIWC can cover a broad spectrum of topic features, it was not specifically designed for health-related topics. LIWC places more emphasis on psychological processes [31,32]. In addition, LIWC was developed almost exclusively in the Western sociocultural context and may not work well when analyzing discussions outside Western societies. During the COVID-19 pandemic, many discussions have been taking place on social media platforms in non–English-speaking regions, such as the Sina Weibo platform in China [33]. Alternative data-driven computational linguistic/NLP algorithms aim to deliver more natural insights directly from data, bypass various human assumptions, overcome lack of inclusiveness of features, and reduce potential bias [34]. Examples of commonly used techniques include word embedding, such as word2vec and doc2vec [35]. However, completely data-driven techniques can result in a lack of interpretability. For instance, the exact meanings of vectors resulting from the doc2vec algorithm are unclear, and it is usually used for classification purposes [25]. Similar to LIWC, it is still challenging to use the Chinese language as an input into these data-driven algorithms without extensive data preprocessing, which may result in a loss of subtlety of the content of the original Chinese post.

Because of these technical challenges, especially the lack of universally designed content analysis and feature extraction analytical workflow, few studies have compared social media discussion across different socio-cultural backgrounds with regard to the COVID-19 pandemic [36-38]. Cross-platform and cross-culture studies are infrequent and generally observational [39,40].

Therefore, we suggest that there is an emergent need to develop a more interpretable and universal content analytical workflow across a wide sociocultural spectrum during the current COVID-19 pandemic and future pandemics. Developing this analytical workflow will vastly expand our fine-grained understanding and characterization of the content features of discussions on health issues worldwide. Until such a workflow is achieved, we will not be able to effectively compare and contrast health communication patterns on different social media platforms worldwide. As such, we propose the following two major objectives in this study:

1. Develop a content feature extraction and coding scheme to characterize discussions about the current COVID-19 pandemic on major social media platforms across socio-cultural backgrounds (Twitter and Sina Weibo);
2. Compare and contrast content features of the most shared viral social media posts on Twitter and Sina Weibo through a comprehensive analytical workflow with state-of-the-art machine learning techniques.

Methods

Retrieval of Social Media Posts

We acquired social media posts on both Sina Weibo (colloquially referred to as Weibo hereafter) and Twitter from January 6 to April 15, 2020, for a total of 100 days. The reasons we used the same sampling period for the two social media platforms were as follows. 1) It made the sampling process consistent and directly comparable; 2) this study focused more on sociocultural than specific geospatial locations. Weibo is almost exclusively used by Chinese users, while Twitter users cover a much wider range of geospatial regions. Given the very different sizes and patterns of the epidemic in different countries, we suggested that having a consistent sampling period could reduce confounding factors such as actual outbreak size and its influence on public perception of COVID-19.

The Weibo posts were acquired via the application programming interface (API) of Hong Kong Baptist University in Python.

We downloaded all Weibo posts during the sampling period without further sampling. There were around 4 million Weibo posts acquired and archived.

The tweets were acquired directly from Twitter via a contract between the School of Data Science, the University of North Carolina at Charlotte, and Twitter. The tweets were not retrieved by the commonly used Twitter API or other commercial APIs. The tweets were a 1% sample; however, given the adequately large sample size (more than 10 million tweets), we believe that this sample is a good representation of public discourse regarding the ongoing pandemic on Twitter.

The keywords used to retrieve social media posts were *COVID19*, *nCOV19*, *SARSCoV2*, their variants (*novel pneumonia*, *SARS*, *SARS2*, *COVID*, *coronavirus*), and other related medical/health terms (*GGO*, *PHEIC*, *pandemic*). Inappropriate, derogative, and discriminating terms such as *WuhanVirus*, *WuhanPenumonia*, and *ChinaVirus* were also included to increase the sample size for research purposes. Both original posts and reposts were retrieved if they included the search terms.

Identification of Viral Posts

“Viral” posts were defined as those with large numbers of shares (also known as “reposts,” “retweets,” etc) on different social media platforms. There are other ways to define viral posts, such as number of comments or number of likes. However, comments may not align with the content and intention of the original posts, while liking would not necessarily propagate the original post on social media. Sharing through reposting or retweeting indicates that the user acknowledged the value of the original post and actively participated in its dissemination on social media. Therefore, the number of shares was used to define viral posts.

Nevertheless, the three types of potential definitions of “viral” post were highly correlated (Pearson correlation coefficient $\rho>0.8$ for each pair of viral post definitions). For example, it was very common for a highly shared COVID-19 post to receive many likes and comments as well. Therefore, we suggest that focusing on one definition of “viral” post was able to provide sufficient insights for the other two definitions.

To avoid oversampling during certain days when a cluster of viral posts occurred (ie, numerous posts occurred on the same day), we identified and selected the 5 most shared posts on Weibo and the 5 most shared posts on twitter throughout the sampling period. Practically, we ordered the posts by original posting date first and then ranked them based on the number of shares they received on each day. Due to the fast pace of social media, most viral posts received a majority of their reposts within a short period of time, and the overall lifespan of a viral post usually lasted less than 48 hours [25]. Eventually, a total of 1000 viral COVID-19 social media posts were selected, 500 for Weibo and 500 for Twitter. Because of the relatively large sample size and size of content feature set (discussed next), we believe the sample size is adequate to provide accurate, granular characterization of viral social media posts regarding COVID-19.

Extraction, Annotation, and Quantification of Content Features

In this study, we developed a relatively novel and comprehensive content analysis workflow to characterize and quantify various content features of health-related social media posts. The creation of content features went through two rounds of iteration. In the first round, we used an open-coding approach to identify an initial set of features by manually analyzing a set of 200 randomly selected social media posts. Then, we randomly selected another set of 800 posts, combined them with the 200 posts, split them into 5 subsets (200 each), and asked five student coders to analyze them independently. The student coders were provided with the list of initial features and were all bilingual, with fluency in both Chinese and English. Moreover, the coders were asked to create new features if they were missing from the existing list. Finally, we refined the list of features based on our review, comparison, and evaluation of the coding results. A few new content features were discussed and added in this round. In the second round, we leveraged the refined features in screening, evaluating, characterizing, and validating a test set of 50 randomly selected posts by our research team. Note that posts in this set were not necessarily viral posts. As discussed later in this paper, randomly selecting posts increases the coverage of various topic contents in the posts. We performed several iterations of intercoder reliability analysis, discussions, and refinements to ensure clarity and consistency in the definition and coding scheme of the features. The intercoder reliability (kappa value) threshold was set as 0.8 before deploying more comprehensive coding. The coding scheme can be described concisely as follows.

Each feature was 0-1 binary coded (ie, a post either had or did not have the specific content feature). This coding scheme is more objective and easier to interpret than LIWC because it only considers the presence of a specific content feature. In addition, because of the objectivity of the coding scheme, there is no need to translate the social media posts, as the subtlety in the original post may be lost during the translation process.

The final version included a total of 77 content features that were grouped into 6 major categories, each major category with more specific features. The six major categories included clinical and epidemiological features (eg, mentioning any symptoms or signs, transmission, or diagnosis and testing); countermeasures and COVID-19–related resources (eg, mentioning face masks, other medical supplies, or disinfection); policies and politics (eg, mentioning social distancing or stay-at-home-orders); public reactions and societal impact (eg, preparedness, remote working, or college education); spatial scales (eg, local, state/provincial, national, or international); and social issues (eg, discrimination against certain countries, violence, uncivil language). Note that these content features were not mutually exclusive, and a post could have multiple features under the same or different major categories at the same time as long as the post contained the specific contents. For example, a single post could mention symptoms, diagnosis, risk factors, and clinical consequences. In addition, these content features were universally developed and objective; therefore, they could be applied in different sociocultural backgrounds without the need of translation, which is required in LIWC. The complete descriptions of these major categories and further specific contents within each major category are provided in Multimedia Appendix 1.

After the comprehensive coding scheme was established and the list of 77 content features was defined, we then coded the 1000 posts according to the coding scheme. For each post, the output was a 77-element 0-1 binary vector. A 1 indicated that the post mentioned the corresponding content feature, while a 0 indicated that the specific post did not mention that content feature. In general, the more 1's (and hence, the fewer 0's), the more diverse the topics contained in the post. Fewer 1's indicated more focused topics in the post. The final output for the analytical workflow was a 1000 × 77 binary matrix that could be further divided into two 500 × 77 binary matrices representing viral Twitter and Sina Weibo groups, respectively.

Descriptive Analysis of Viral COVID-19 Posts Across Social Media Platforms

We applied descriptive analysis to quantify and contrast the prevalence of content features in the most viral COVID-19 posts across the social media platforms Weibo and Twitter. The prevalence was defined as the percentage of number of 1's across all the sampled posts in each content feature. Prevalence was bounded between 0 (ie, none of the sampled posts mentioned the content feature) and 1 (ie, all posts mentioned the content feature). A larger prevalence indicated that the corresponding content feature was more frequently mentioned in the viral social media posts regarding COVID-19.

We further applied a two-sample *z* test to investigate whether there was statistically significant differences in the two prevalence measures of the same content feature between Weibo and Twitter. Because the data were 0-1 binary instead of continuous, the *z* test was more appropriate than the *t* test or Kolmogorov-Smirnov test. The content features that had the most distinct prevalence measures between the two social media platforms were identified based on the *z* test.

In addition to comparing different social media posts, we also studied the associations between content features on different social media platforms. Pairwise Pearson correlation was calculated between each pair of content features in both Twitter and Sina Weibo posts. Pairs with statistically significant associations ($P<.05$) were identified. These analyses provide a comprehensive characterization on how viral COVID-19 content features are distributed and correlated differently on the two major social media platforms.

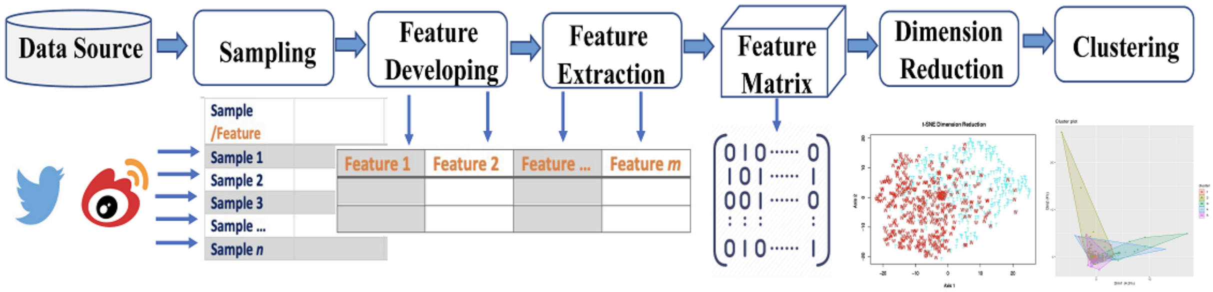
Unsupervised Learning of Viral COVID-19 Posts Across Social Media Platforms

To further investigate the distributions and relationships among multiple content features simultaneously, we applied the *t* distribution stochastic neighbor embedding (*t*-SNE) technique. *t*-SNE is a machine learning dimension reduction algorithm. In contrast to the more commonly used principal component analysis technique, *t*-SNE can handle data that are not normally distributed, as presented in this study (ie, binary data) and is also commonly used in other studies involving large and heterogeneous data (eg, bioinformatics data [41]). Performing *t*-SNE provides a clear visualization of associations among content features in 2D space instead of the original complex 77-dimensional feature space.

t-SNE dimension reduction paved the way for subsequent clustering analysis. In this study, we applied unsupervised machine learning k-means clustering [42]. Note that we created 6 major categories of content features for our own manual content coding effort. These 6 categories were based on our observation and discussion about the COVID-19 pandemic and public discourse on social media. Data-driven clustering analysis (also known as unsupervised learning), on the other hand, enables the data to "speak for themselves" (hence, "unsupervised"). Data-driven clustering provides a new angle of identifying possible aggregations of content features. For example, frequently concurrent content features may not necessarily be clustered under the same major manually created categories. *k*-means clustering does not require a priori information from researchers on how the features should be grouped; therefore, it reduces potential bias. The optimal *k* value to perform *k*-means clustering was determined by computing and inspecting the total within sum of squares (TWSS) with a wide range of *k* values from 1 to 20. Although larger *k* values are usually associated with smaller TWSSs, they increase the difficulty of interpreting the clusters. We examined and contrasted the clustering patterns of content features in the most viral COVID-19 posts on Twitter and Weibo.

The complete workflow of extracting and analyzing viral COVID-19 posts on different social media platforms is conceptualized and presented in Figure 1. All analytical codes were developed in R 4.0.2 (R Project) with supporting packages of *Rtsne*, *tidyverse*, *cluster*, *factoextra*, *gridExtra*, *wordcloud*, *tm*, *corrplot*, and *ggplot2*. The codes and data are freely available upon request.

Figure 1. Conceptualized social media content feature extraction and analysis workflow. Sampling was performed with specific application programming interfaces in Python. Feature developing, extraction, and quantification were performed manually by our group. Subsequent analyses were performed in R.

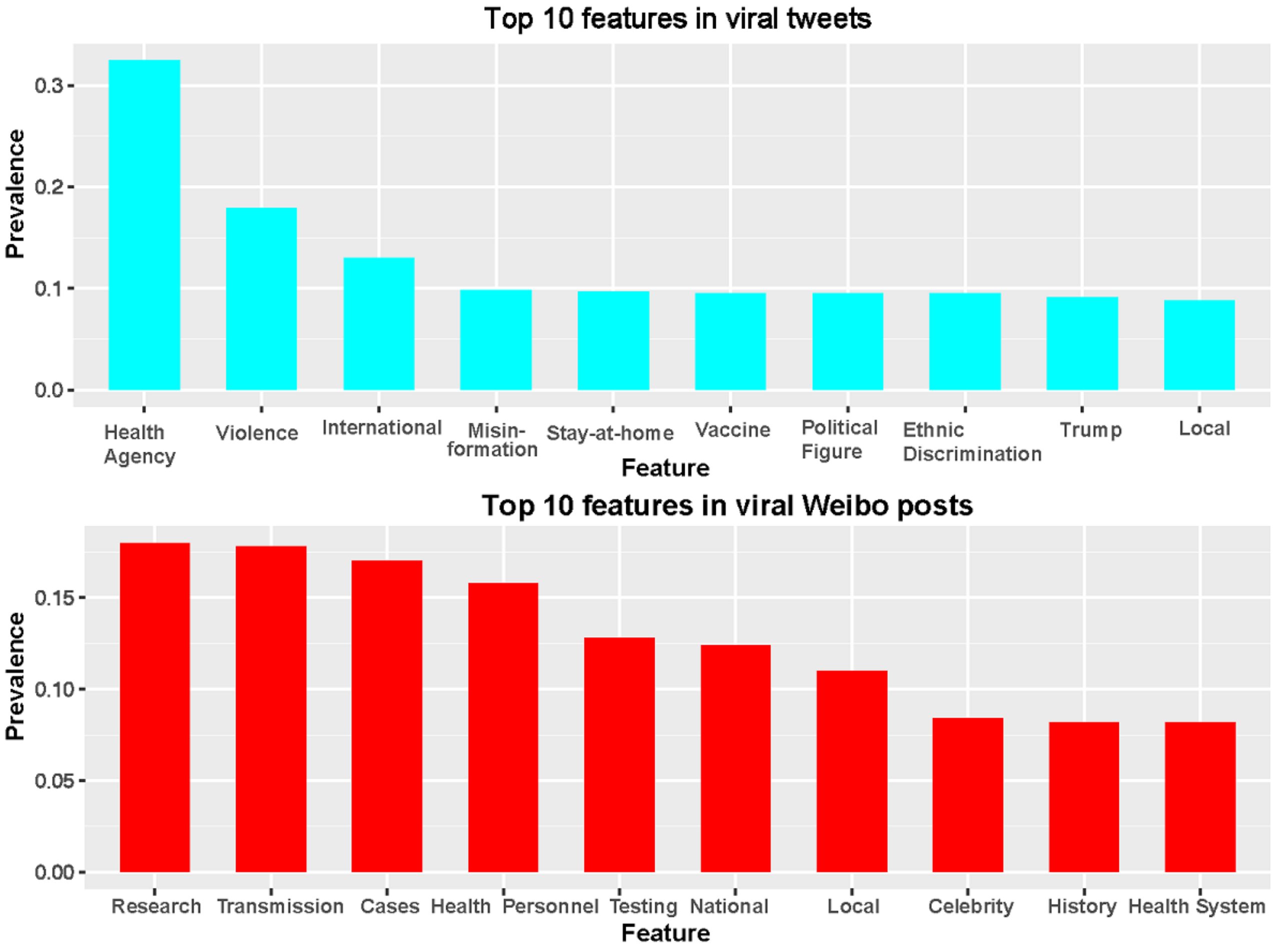


Results

Description of Viral COVID-19–Related Social Media Posts on Sina Weibo and Twitter

The most prevalent content features in Twitter (which has mostly Western users) were *health agency* (eg, CDC [US Centers for Disease Control and Prevention], NIH [National Institutes of Health], and WHO [World Health Organization]; 37.0%), *violence* (mostly related to domestic violence due to stay-at-home orders; 20.4%), *international relationships* (14.8%), *misinformation* (eg, mentioning *misinformation*, *disinformation*, *hoax*, *fake news*; 11.2%), *stay-at-home order* (11.0%), and *vaccine* (10.8%). The 10 most frequently mentioned content features on Twitter, along with their prevalence and ranking, are shown in Figure 2 (top panel). In general, prevalent COVID-19 content features on Twitter did not directly focus on the disease itself and the epidemic but rather on policies, politics, and other secondary societal issues, such as violence and discrimination. This finding reinforced the notion that COVID-19, like many large pandemics and emerging health issues, is not an isolated medical issue and is intertwined with complicated sociopolitical aspects. In particular, 2020 was a US presidential election year. Therefore, it was not surprising that US President Donald Trump and other former and current US office holders (eg, President Barack Obama, Vice President Joe Biden, Majority Leader Mitch McConnell, and House Speaker Nancy Pelosi) were frequently mentioned in COVID-19–related viral tweets. Given the partisan nature of the US political system, the *Republican Party* and *Democratic Party* were also consistently mentioned with COVID-19, mostly with the distinct views and countermeasures of these parties related to the pandemic. The most mentioned nonpolitician *celebrity* was Bill Gates, and mentions of his name were usually associated with content features of *vaccines* and *misinformation* (mostly vaccine-related conspiracy theories). *Discrimination* toward Chinese people, Asian Americans, and Asian people in general was also frequently mentioned. Note that these were content features and may not reflect actual discrimination and negative sentiments against these groups in the tweets. In fact, many viral tweets that mentioned *discrimination* features were advocating for the elimination of discrimination and xenophobia.

Figure 2. Top 10 content features and their prevalence on Twitter (top) and Weibo (bottom).



In comparison, the most prevalent content features in Weibo were *research* (18%), *transmission* (17.8%), *cases* (17%), *healthcare personnel* (15.8%), and *testing* (12.8%). The top 10 most mentioned content features on Weibo, along with their prevalence and ranking, are shown in Figure 2 (bottom panel). Compared to Twitter users, Weibo users (who are mostly Chinese) were more likely to engage in discussion of disease-related content features. Among the 10 most common content features, only *celebrity* was not directly related to the disease itself. In other words, Chinese Weibo users tended to focus on COVID-19 as a health and medical issue rather than on the associated societal and political issues discussed in Western societies. Viral Weibo posts were much more likely to mention *health personnel* and pay tribute to health care workers. *Research* on the SARS-CoV-2 pathogen and its *transmission* among human populations were also frequently mentioned, demonstrating the public interest in the state-of-the-art understanding of the emerging health crisis. Because China experienced the original 2003 severe acute respiratory syndrome (SARS) outbreak, which was caused by severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), and COVID-19 was caused by a similar coronavirus (SARS-CoV-2), the *history* of the 2003 SARS outbreak was a recurrent theme in COVID-19 Weibo posts. The *celebrities* mentioned in posts related to COVID-19 on Weibo were also very different from those on Twitter. In general, viral Weibo posts mentioned pop culture idols (eg, singers, other performing artists, and sports stars), and the sentiment was almost always positive (eg, mentions of financial, resource, and emotional support for COVID-19–impacted regions and people provided by these celebrities).

These results showed vastly different content features covered in viral posts between Weibo and Twitter, which reflected the vast differences in perception of COVID-19 in the corresponding two major sociocultural systems. In general, Twitter users (who mostly live in Western countries) were highly engaged in discussions with countermeasures, politics, and policies related to the COVID-19 pandemic. In comparison, Weibo users (mostly Chinese) tended to focus more on the disease itself, but not exclusively. Among the top 10 features, the only overlapping content feature between the two platforms was the *local* situation. Therefore, these findings reveal substantially different focuses on the COVID-19 pandemic in Chinese and Western societies, which were reflected in the most viral social media posts in cyberspace.

Comparative Analysis of Content Features of Twitter and Sina Weibo

We further provided a quantitative comparison of content features between the two social media platforms. Out of a total of 77 content features, 3 (4%) were absent from all of the 500 most viral tweets (*comorbidity*, *eHealth*, and *suicide*), and 6 (8%) were not present in any of the 500 most viral Weibo posts (*constitution*, *curfew*, *remote working*, *major religion*, *discrimination against gender*, and *discrimination against religion*). This result also implies that viral discussions of COVID-19 on Weibo had narrower but more focused content features. There was no intersection of missing features between the two major social media platforms.

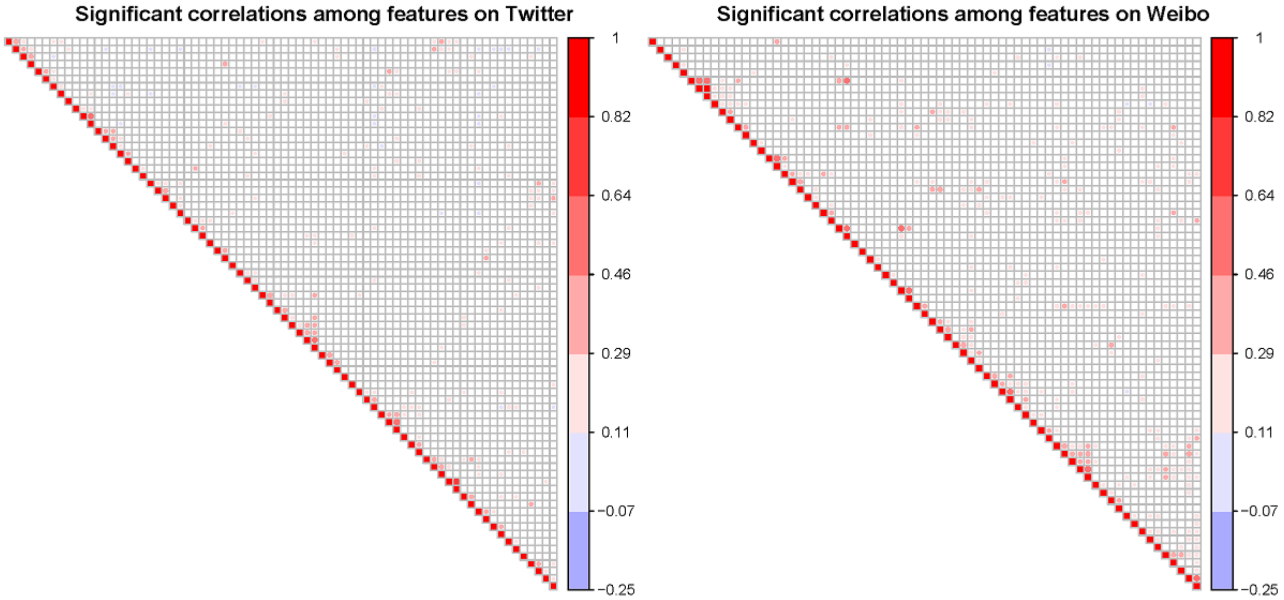
Two-sample z tests were used to further quantify between-platform differences for each content feature. Content features with zero prevalence (ie, never mentioned in viral social media posts on either platform) were removed to perform the z test correctly. Features having the most distinct prevalence between the two platforms were *health agency* (difference of prevalence [D]=0.25; Twitter minus Weibo; $P<.001$), *vaccine* ($D=–0.17$, $P<.001$), *shelter-in-place* (or lockdown, $D=–0.11$, $P<.001$), *cases* ($D=0.09$, $P<.001$), and *stay-at-home order* ($D=0.10$, $P=.002$). While many of these content features were among the top 20 mentioned on both social media platforms (Figure 2), we also observed that *local* situations, the only common top 10 feature in both platforms, actually had statistically significant differences ($D=–0.11$, $P<.001$). *Local* was the 6th most mentioned content feature on Weibo and the 10th on Twitter. These quantitative findings can be explained by the different sociocultural backgrounds of the users of Twitter (Western) and Weibo (Chinese).

Some features were also distributed similarly between the two social media platforms (ie, P values substantially greater than .05 based on the z test). Of them, *preparedness* ($D<0.01$, $P=.90$), *discrimination against ethnicity* ($D<0.01$, $P=.96$), *prevention* ($D<0.01$, $P=.97$), *recovery* ($D<0.01$, $P=.97$), *ecosystem* ($D<0.01$, $P=.97$), *masks* ($D<0.01$, $P>.99$), and *Trump* ($D<0.01$, $P>.99$) were the least distinct features. These features represent the common ground regarding COVID-19 between the two social media platforms and the two underlying sociocultural systems.

The missing content features revealed a discrepancy between viral and nonviral discussions of COVID-19 on social media. As mentioned earlier, the comprehensive content feature coding scheme was originally developed from a random sample of posts, most of which were nonviral posts with <5 reposts. We speculated that certain controversial content features (especially those related to policy and politics on Twitter) facilitated the spread of certain posts on social media and caused them to go viral. Posts that are less controversial typically do not gain much attention and do not go viral on social media. However, we must point out that content features are only one reason that a post can go viral. Other aspects include temporality (ie, when the post was published relative to the epidemic), property of the original posting user (eg, number of followers), and the severity of the pandemic at that time and place.

Significant Pearson correlations ($P<.05$) are shown in Figure 3 for Twitter (left) and Weibo (right) posts, respectively. In general, significantly correlated content feature pairs were more abundant on Weibo than on Twitter. One possible explanation is that Twitter has a 280-character length limit for posts. Therefore, content features in each tweet were limited, and concurrent content features in the same tweet were less frequent. On the other hand, Sina Weibo allows up to 2000 characters; therefore, it is possible to include much more content in a Weibo post than in a tweet. Consequently, a Weibo post can accommodate more content features than a tweet. Viral COVID-19 tweets included an average of 2.37 content features, and viral Weibo posts contained 2.78 content features. However, most viral Weibo posts used URLs to pack in more information and keep the post concise rather than including everything in the main post content. Therefore, the 2000 character limit is only a theoretical upper limit and was rarely reached, especially for viral Weibo posts.

Figure 3. Significant Pearson correlations ($P<.05$) among content features on Twitter (left) and Weibo (right). The complete list of the 77 content features on the x- and y-axes can be found in Multimedia Appendix 1.

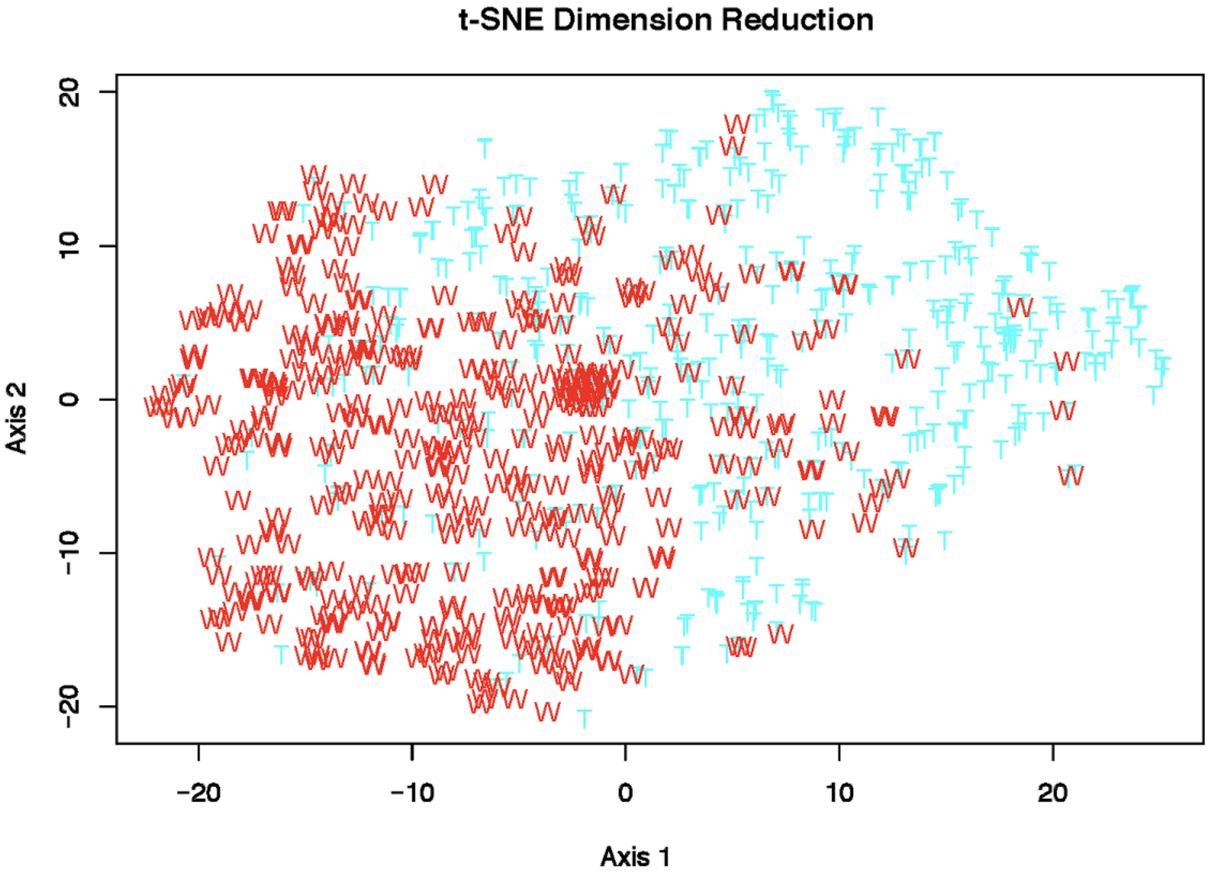


Note that Weibo is subject to censorship toward certain content features. For example, although US President *Trump* was mentioned quite a few times in viral Weibo posts, President Xi of China is not an allowed topic on Weibo and Chinese cyberspace in general. Therefore, there was no equivalent content feature to *Trump* on Weibo. Other *political figures* in China, such as the governor of Hubei (Yong Ying), are generally permitted by censors to be mentioned and commented on in Weibo posts.

Dimension Reduction and Clustering Analysis of Content Features

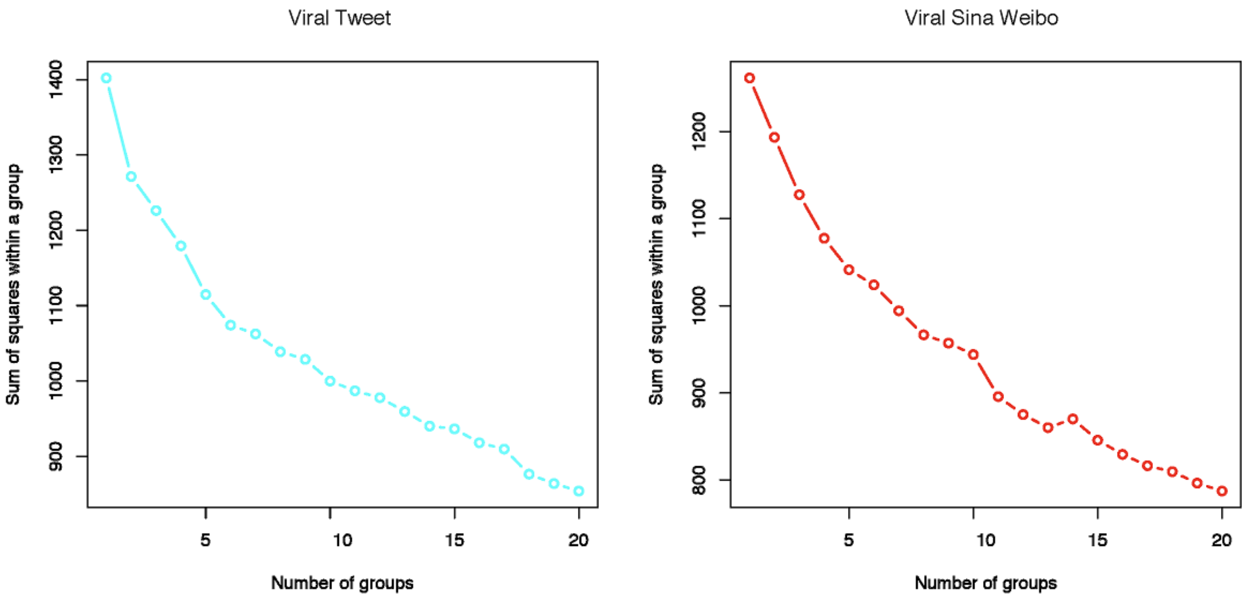
The machine learning dimension reduction *t*-SNE results for Twitter and Weibo are shown in Figure 4. These figures show how content features are distributed and associated in the reduced 2D space instead of the original 77-dimensional feature space. It is very clear that the content features have distinct distribution patterns between the two social media platforms in the reduced 2D space. This reinforces our previous findings on the variability of content features across the sociocultural spectrum.

Figure 4. *t*-SNE results for viral COVID-19 tweets (T) and Weibo posts (W).



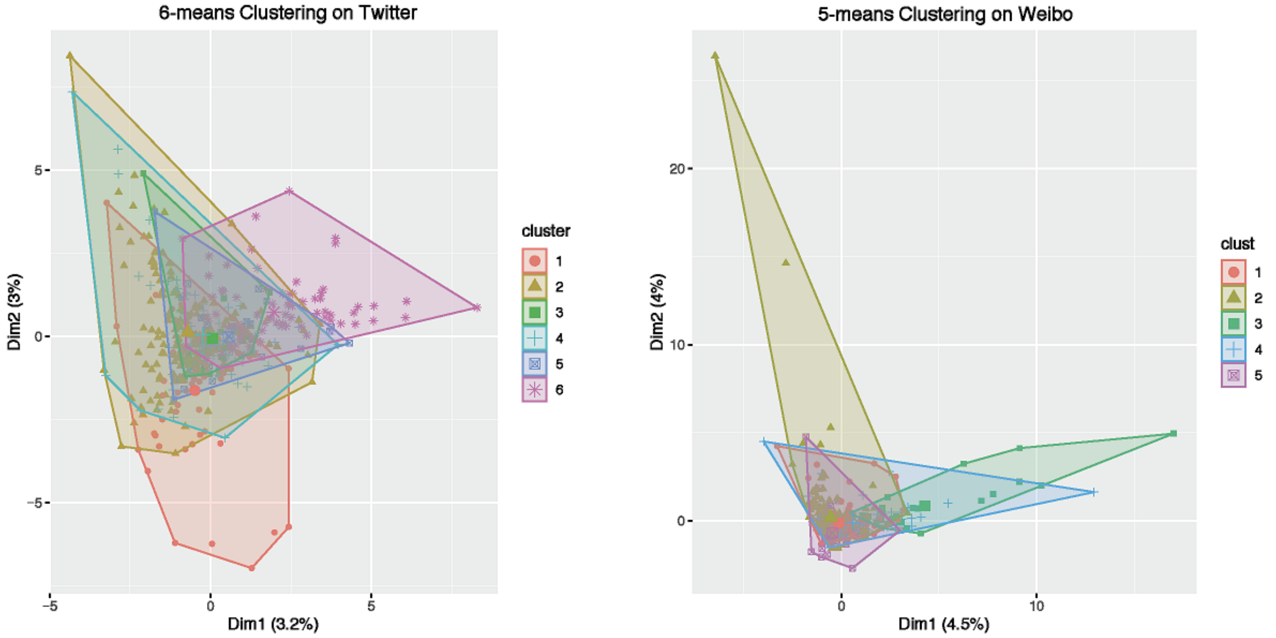
The number of optimal clusters on Twitter (k_t) was determined as 6 from Figure 5 (left), while the number of optimal clusters (k_s) on Weibo was found to be 5 from Figure 5 (right). Therefore, not only were content features regarding COVID-19 distributed differently between the two social media platforms, but their associations (eg, clusters) within posts were also distinct between the two platforms. Note that these clusters were identified by the data-driven unsupervised machine learning technique, and these clusters did not necessarily align with the 6 manually developed major categories.

Figure 5. Numbers of clusters and within-group sums of squares for tweets (left) and Weibo posts (right).



We further show the *k*-means clustering results of the content features on Twitter and Weibo in Figure 6 (left and right, respectively). The clustering patterns were substantially different between the two social media platforms. The sizes of the 6 distinct clusters on Twitter were 154, 107, 96, 62, 42, and 39. The total sum of squares (TSS) across all 6 clusters was 1402. The total within-cluster sum of squares (TWSS) was 1079, and the total between-cluster sum of squares (TBSS) was 323 on Twitter. Note that TSS = TWSS + TBSS. In comparison, the 5 cluster sizes of Weibo posts were 218, 106, 81, 67, and 28. The TSS, TWSS, and TBSS on Weibo were 1262, 1034, and 228, respectively. Therefore, all sums of squares were much smaller on Weibo than on Twitter. In addition, the two dimensions (the x- and y-axes in Figure 6) were also much smaller on Twitter (3.2% and 3%) than on Weibo (4.5% and 4%). All these results reveal that COVID-19 content features in viral Weibo posts were more similar across different posts than those in Twitter posts. Twitter showed a more diverse array of content features among different tweets.

Figure 6. *k*-means clustering on viral tweets (left) and Weibo posts (right).



Discussion

Theoretical Innovation

This study is the first of its kind to comprehensively characterize the content features of discussions regarding a large pandemic on social media across the sociocultural spectrum. We showed the vast differences in topic content features of viral social media posts in Twitter and Weibo, the two most influential social media platforms in China and the West during the COVID-19 pandemic. In general, viral social media posts in China focused on cases and prevention, which are topics that are more related to COVID-19 as a health issue. However, as a comparison, most viral tweets regarding COVID-19 focused more on policies and politics, including *stay-at-home* orders, President *Trump*, and other *political figures*. Through various analytical methods, social media data provided a new angle to explore and understand public discourse of the COVID-19 pandemic and associated social, political, and economic issues. Details of these discussions in virtual cyberspace may provide insights on the actual disease epidemic in the real world. For example, analyzing public perception of various NPIs, such as *social distancing*, mandatory *mask-wearing*, and *stay-at-home* orders, would provide an estimation of the compliance with these NPIs, which determine the case counts and epidemic trajectory in a region. This concept echoes the original idea of infodemiology, which uses a time series of social media post counts related to a health issue (eg, COVID-19) as an indicator of actual case counts [14-16]. In addition to the number of social media posts, we will be able to further extract fine-grained perceptions of the risk and NPIs of COVID-19 and extend the application of infodemiology.

Technical Advances

To achieve an effective comparison across the sociocultural spectrum regarding the COVID-19 pandemic on social media, we developed a comprehensive content analytical workflow. This analytical workflow was specifically designed for transboundary infectious diseases (eg, outbreaks and pandemics of infectious diseases) that have complicated sociocultural contexts. Compared to the commonly used LIWC [31], our workflow, especially the coding scheme, has several advantages. First, LIWC is a general content analytical tool that ignores many important content features during the COVID-19 pandemic. Our coding scheme is tailored to the complicated and interacting health, social, cultural, and political nuances of transboundary infectious diseases. Therefore, our coding scheme is able to capture a much more comprehensive and detailed content features in web-based discussions regarding transboundary infectious diseases. Second, LIWC uses proprietary algorithms to calculate individual scores of different features, and the exact interpretation of the numeric values is not readily comprehensible. In contrast, our coding scheme is 0-1 binary, where 1 indicates that the content has a feature and 0 indicates that it does not. This coding is clearer than the obscure LIWC scores. In addition, LIWC scores vary substantially (from 0 to 100) among different features. Certain features that have large values in LIWC tend to dominate and overshadow other features; thus, further analysis is prone to bias. Our coding scheme is consistent, as all features have the same coding scheme. Finally, LIWC is difficult to directly apply to non–Indo-European languages; therefore, direct comparison between sociocultural contexts with LIWC is almost impossible. In contrast, our coding scheme is context-free and can be applied to virtually any language and any region. The coding scheme itself is also flexible. Researchers can easily add and modify content features as necessary when working with other health issues beyond COVID-19. The coding scheme can be retrofitted to understand communications on previous events (eg, the 2016 Zika event). We can easily add, remove, or revise corresponding content features related to the specific health issues we are exploring.

Limitations of the Current Study and Future Directions

This study adopts a static view of all viral social media posts for comparative analysis between two sides of the sociocultural spectrum in a given period of time. However, for a large and ongoing pandemic, time is another major influential factor that is associated with the actual progress of the pandemic. Our previous studies showed that the Zika case series was strongly associated with the Zika discussion trend on Twitter in 2016 [24,25]. Similarly, future studies can be expended to explicitly characterize how various content features evolve with time in different regions. The ongoing COVID-19 pandemic case series can be predicted by certain content features (eg, regarding NPIs), similar to the previously discussed infodemiology approach.

We used the number of reposts (ie, retweets or shares) as the definition of a viral social media post. One limitation is that we did not consider the possibility of automatic reposting by bots or cyborgs. Therefore, it is possible that the large number of reposts may not accurately represent and reflect the public perception of an issue. Bots and cyborgs, however, are not necessarily associated with misinformation. Bots and cyborgs can be used as tools to quickly disseminate information on social media platforms for other reasons, such as advertising. A future direction of this study is to identify other definitions of viral posts (eg, posts with a large number of likes, favorites, or comments).

Viral social media posts are only one of many attributes of social media discussion. Our initial assessment showed that >75% of tweets and >80% of Weibo posts regarding COVID-19 did not receive any attention on social media. This number is similar to our previous finding that 76% of all Zika-related tweets were never retweeted [25].

To characterize web-based public discourse related to COVID-19 and other emerging health issues accurately and comprehensively, we will continue studying these nonviral social media posts on different platforms. However, given the ever-increasing volume of social media posts, effective sampling strategies are a priority. Effective sampling is a necessity to provide a less biased depiction of content features. Data mining of nonviral posts regarding COVID-19, especially on sentiment toward NPIs, will provide a more accurate estimation of compliance with NPIs in different regions at different stages of the pandemic. We will also be able to further compare and contrast how the distributions of content features differ between viral and nonviral post groups as well as across the sociocultural spectrum.

In this study, we depict how NPIs of COVID-19 have been mentioned on social media across the sociocultural spectrum. Because this study focuses on providing a neutral and objective characterization of content features in COVID-19–related discussions, it does not consider subjective sentiment toward specific NPIs. However, individual and societal perception toward NPIs can be strong influencing factors during the COVID-19 pandemic. For instance, positive sentiment toward *mask-wearing* and *social distancing* may reflect actual compliance with these NPIs in society and hence help reduce the risk of transmission. On the other hand, negative sentiment toward these NPIs may lead to noncompliance and facilitate COVID-19 transmission in the real world. In a future study, we will further integrate objective content features and corresponding sentiment and/or emotion to provide a more comprehensive understanding of public perceptions.

Finally, this study relies on human coding of content features, which is substantially labor-intensive. For instance, adequate and proper training is required to achieve high intercoder reliability before each coder can perform independently. In comparison, the LIWC algorithm is automated and relatively easy to use. We are still at the early development stage of a novel analytical workflow that is similar to LIWC. We expect to develop at least a semiautomated and semisupervised machine learning method for quick and effective web-based health information processing and annotation. To achieve this ambitious goal, we envision a crowd-sourcing approach that will enable ardent citizen scientists and volunteers worldwide to help further manually code more social media posts, create an even larger corpus, and develop state-of-the-art semisupervised or supervised machine learning pipelines to automate the process. The eventual product will be able to automatically extract content features from social media posts regarding health issues and can further guide effective health communications during emergencies.

Acknowledgments

SC is supported by the Models of Infectious Disease Agent Study (MIDAS) COVID-19 urgent supplementary grant (MIDASUP2020-5). YS is supported by the Interdisciplinary Research Clusters Matching Scheme (IRCMS/19-20/D04) and the AI and Media Research Lab at Hong Kong Baptist University (SDF17-1013-P01). The authors are grateful for the help from Ms Mengyu Li and Mr Minghao Wang from Hong Kong Baptist University and Mr Evan Lai from St. Mark's School of Texas for their participation in and contributions to this project. The authors are also grateful for the technical assistance from the School of Data Science, UNC Charlotte, in providing the social media data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coding book of COVID-19 content features.
[PDF File (Adobe PDF File), 155 KB - jmir_v23i1e24889_app1.pdf]

References

1. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
2. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
3. Eckert S, Sopory P, Day A, Wilkins L, Padgett D, Novak J, et al. Health-Related Disaster Communication and Social Media: Mixed-Method Systematic Review. Health Commun 2018 Dec;33(12):1389-1400 [FREE Full text] [doi: 10.1080/10410236.2017.1351278] [Medline: 28825501]
4. Grajales F, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
5. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]
6. Hughes A, Palen L. The evolving role of the public information officer: An examination of social media in emergency management. J Homel Secur Emerg Manag 2012;9(1):22. [doi: 10.1515/1547-7355.1976]
7. Wiyeh A, Cooper S, Jaca A, Mavundza E, Ndwandwe D, Wiysonge CS. Social media and HPV vaccination: Unsolicited public comments on a Facebook post by the Western Cape Department of Health provide insights into determinants of vaccine hesitancy in South Africa. Vaccine 2019 Oct 08;37(43):6317-6323 [FREE Full text] [doi: 10.1016/j.vaccine.2019.09.019] [Medline: 31521412]

8. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/ajph.2018.304567]
9. Naslund J, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
10. Beall AT, Hofer MK, Schaller M. Infections and Elections: Did an Ebola Outbreak Influence the 2014 U.S. Federal Elections (and if so, How)? Psychol Sci 2016 May 14;27(5):595-605. [doi: 10.1177/0956797616628861] [Medline: 26976083]
11. Davies SE, Bennett B. A gendered human rights analysis of Ebola and Zika: locating gender in global health emergencies. Int Aff 2016 Aug 31;92(5):1041-1060. [doi: 10.1111/1468-2346.12704]
12. Adebayo G, Neumark Y, Gesser-Edelsburg A, Abu AW, Levine H. Zika pandemic online trends, incidence and health risk communication: a time trend study. BMJ Glob Health 2017 Aug;2(3):e000296 [FREE Full text] [doi: 10.1136/bmjgh-2017-000296] [Medline: 29082006]
13. Avery EJ. Public information officers' social media monitoring during the Zika virus crisis, a global health threat surrounded by public uncertainty. Public Relat Rev 2017 Sep;43(3):468-476. [doi: 10.1016/j.pubrev.2017.02.018]
14. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002 Dec;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0]
15. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
16. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158 [FREE Full text] [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
17. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
18. Brenner CJ, Inbar Y. Disgust sensitivity predicts political ideology and policy attitudes in the Netherlands. Eur J Soc Psychol 2014 Nov 15;45(1):27-38. [doi: 10.1002/ejsp.2072]
19. Faulkner J, Schaller M, Park JH, Duncan LA. Evolved Disease-Avoidance Mechanisms and Contemporary Xenophobic Attitudes. Group Process Intergroup Relat 2004 Sep 20;7(4):333-353. [doi: 10.1177/1368430204046142]
20. Fernández-Luque L, Bau T. Health and social media: perfect storm of information. Healthc Inform Res 2015 Apr;21(2):67-73 [FREE Full text] [doi: 10.4258/hir.2015.21.2.67] [Medline: 25995958]
21. Kou Y, Gui X, Chen Y, Pine K. Conspiracy Talk on Social Media. Proc ACM Hum-Comput Interact 2017 Dec 06;1(CSCW):1-21. [doi: 10.1145/3134696]
22. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014 Oct 14;349:g6178. [doi: 10.1136/bmj.g6178] [Medline: 25315514]
23. Shadgan B, Pakravan A, Zaeimkohan H, Shahpar FM, Khodaee M. Zika and Rio Olympic Games. Curr Sports Med Rep 2016;15(4):298-300. [doi: 10.1249/jsr.0000000000000278]
24. Chen S, Xu Q, Buchenberger J, Bagavathi A, Fair G, Shaikh S, et al. Dynamics of Health Agency Response and Public Engagement in Public Health Emergency: A Case Study of CDC Tweeting Patterns During the 2016 Zika Epidemic. JMIR Public Health Surveill 2018 Nov 22;4(4):e10827 [FREE Full text] [doi: 10.2196/10827] [Medline: 30467106]
25. Safarnejad L, Xu Q, Ge Y, Bagavathi A, Krishnan S, Chen S. Identifying Influential Factors in the Discussion Dynamics of Emerging Health Issues on Social Media: Computational Study. JMIR Public Health Surveill 2020 Jul 28;6(3):e17175 [FREE Full text] [doi: 10.2196/17175] [Medline: 32348275]
26. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
27. Ataguba OA, Ataguba JE. Social determinants of health: the role of effective communication in the COVID-19 pandemic in developing countries. Glob Health Action 2020 Dec 31;13(1):1788263 [FREE Full text] [doi: 10.1080/16549716.2020.1788263] [Medline: 32657669]
28. Lwin M, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
29. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
30. Pennebaker J, Booth RJ, Boyd RL, Francis ME. Linguistic Inquiry and Word Count: LIWC2015. Pennebaker Conglomerates. 2015. URL: https://s3-us-west-2.amazonaws.com/downloads.liwc.net/LIWC2015_OperatorManual.pdf [accessed 2020-12-18]

31. Zhao N, Jiao D, Bai S, Zhu T. Evaluating the Validity of Simplified Chinese Version of LIWC in Detecting Psychological Expressions in Short Texts on Social Network Services. PLoS One 2016;11(6):e0157947 [FREE Full text] [doi: 10.1371/journal.pone.0157947] [Medline: 27322382]
32. Guntuku S, Schneider R, Pelullo A, Young J, Wong V, Ungar L, et al. Studying expressions of loneliness in individuals using twitter: an observational study. BMJ Open 2019 Nov 04;9(11):e030355 [FREE Full text] [doi: 10.1136/bmjopen-2019-030355] [Medline: 31685502]
33. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
34. Turner CA, Jacobs AD, Marques CK, Oates JC, Kamen DL, Anderson PE, et al. Word2Vec inversion and traditional text classifiers for phenotyping lupus. BMC Med Inform Decis Mak 2017 Aug 22;17(1):126 [FREE Full text] [doi: 10.1186/s12911-017-0518-1] [Medline: 28830409]
35. Mikolov T. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. 2013 Dec Presented at: 26th International Conference on Neural Information Processing Systems; December 5-10, 2013; Lake Tahoe, NV p. 3111-3119 URL: https://dl.acm.org/doi/10.5555/2999792.2999959
36. Dkhar S, Quansar R, Saleem S, Khan S. Indian J Public Health 2020;64(6):205 [FREE Full text] [doi: 10.4103/ijph.ijph_469_20]
37. Hussain I, Majeed A, Imran I, Ullah M, Hashmi FK, Saeed H, et al. Knowledge, Attitude, and Practices Toward COVID-19 in Primary Healthcare Providers: A Cross-Sectional Study from Three Tertiary Care Hospitals of Peshawar, Pakistan. J Community Health 2020 Jul 06 [FREE Full text] [doi: 10.1007/s10900-020-00879-9] [Medline: 32632645]
38. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
39. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-Country Comparison of Public Awareness, Rumors, and Behavioral Responses to the COVID-19 Epidemic: Infodemiology Study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
40. Tao Z, Chu G, McGrath C, Hua F, Leung YY, Yang W, et al. Nature and Diffusion of COVID-19-related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo. J Med Internet Res 2020 Jun 15;22(6):e19981 [FREE Full text] [doi: 10.2196/19981] [Medline: 32501808]
41. Li W, Cerise JE, Yang Y, Han H. Application of t-SNE to human genetic data. J Bioinform Comput Biol 2017 Aug 22;15(04):1750017. [doi: 10.1142/s0219720017500172]
42. Demidenko E. The next-generation K-means algorithm. Stat Anal Data Min 2018 Aug;11(4):153-166 [FREE Full text] [doi: 10.1002/sam.11379] [Medline: 30073045]

Abbreviations

CDC: US Centers for Disease Control and Prevention
***D*:** difference of prevalence
LIWC: linguistic inquiry and word count
NIH: National Institutes of Health
NLP: natural language processing
NPI: nonpharmaceutical interventions
PHEIC: public health emergency of international concern
SARS: severe acute respiratory syndrome
SARS-CoV-1: severe acute respiratory syndrome coronavirus 1
***t*-SNE:** t distribution stochastic neighbor embedding
TBSS: total between-cluster sum of squares
TSS: total sum of squares
TWSS: total within sum of squares
WHO: World Health Organization

Edited by G Eysenbach; submitted 12.10.20; peer-reviewed by F Yu, R Sun, R Zowalla; comments to author 03.11.20; revised version received 24.11.20; accepted 03.12.20; published 06.01.21.

Please cite as:
Chen S, Zhou L, Song Y, Xu Q, Wang P, Wang K, Ge Y, Janies D
A Novel Machine Learning Framework for Comparison of Viral COVID-19–Related Sina Weibo and Twitter Posts: Workflow Development and Content Analysis
J Med Internet Res 2021;23(1):e24889
URL: https://www.jmir.org/2021/1/e24889
doi:10.2196/24889
PMID:33326408

Review

Prevalence of Health Misinformation on Social Media: Systematic Review

Victor Suarez-Lledo[1,2*], BSc, MSc; Javier Alvarez-Galvez[1,2*], BSc, MSc, PhD

[1]Department of Biomedicine, Biotechnology and Public Health, University of Cadiz, Cadiz, Spain

[2]Computational Social Science DataLab, University Research Institute on Social Sciences, University of Cadiz, Jerez de la Frontera, Cadiz, Spain

[*]all authors contributed equally

Corresponding Author:
Victor Suarez-Lledo, BSc, MSc
Department of Biomedicine, Biotechnology and Public Health
University of Cadiz
Av Ana de Viya, 52
Cadiz, 11009
Spain
Phone: 34 956019080
Email: victor.sanz@uca.es

Abstract

Background: Although at present there is broad agreement among researchers, health professionals, and policy makers on the need to control and combat health misinformation, the magnitude of this problem is still unknown. Consequently, it is fundamental to discover both the most prevalent health topics and the social media platforms from which these topics are initially framed and subsequently disseminated.

Objective: This systematic review aimed to identify the main health misinformation topics and their prevalence on different social media platforms, focusing on methodological quality and the diverse solutions that are being implemented to address this public health concern.

Methods: We searched PubMed, MEDLINE, Scopus, and Web of Science for articles published in English before March 2019, with a focus on the study of health misinformation in social media. We defined health misinformation as a health-related claim that is based on anecdotal evidence, false, or misleading owing to the lack of existing scientific knowledge. We included (1) articles that focused on health misinformation in social media, including those in which the authors discussed the consequences or purposes of health misinformation and (2) studies that described empirical findings regarding the measurement of health misinformation on these platforms.

Results: A total of 69 studies were identified as eligible, and they covered a wide range of health topics and social media platforms. The topics were articulated around the following six principal categories: vaccines (32%), drugs or smoking (22%), noncommunicable diseases (19%), pandemics (10%), eating disorders (9%), and medical treatments (7%). Studies were mainly based on the following five methodological approaches: social network analysis (28%), evaluating content (26%), evaluating quality (24%), content/text analysis (16%), and sentiment analysis (6%). Health misinformation was most prevalent in studies related to smoking products and drugs such as opioids and marijuana. Posts with misinformation reached 87% in some studies. Health misinformation about vaccines was also very common (43%), with the human papilloma virus vaccine being the most affected. Health misinformation related to diets or pro–eating disorder arguments were moderate in comparison to the aforementioned topics (36%). Studies focused on diseases (ie, noncommunicable diseases and pandemics) also reported moderate misinformation rates (40%), especially in the case of cancer. Finally, the lowest levels of health misinformation were related to medical treatments (30%).

Conclusions: The prevalence of health misinformation was the highest on Twitter and on issues related to smoking products and drugs. However, misinformation on major public health issues, such as vaccines and diseases, was also high. Our study offers a comprehensive characterization of the dominant health misinformation topics and a comprehensive description of their prevalence on different social media platforms, which can guide future studies and help in the development of evidence-based digital policy action plans.

(J Med Internet Res 2021;23(1):e17187) doi:10.2196/17187

KEYWORDS

social media; health misinformation; infodemiology; infodemics; social networks; poor quality information; social contagion

Introduction

Over the last two decades, internet users have been increasingly using social media to seek and share health information [1]. These social platforms have gained wider participation among health information consumers from all social groups regardless of gender or age [2]. Health professionals and organizations are also using this medium to disseminate health-related knowledge on healthy habits and medical information for disease prevention, as it represents an unprecedented opportunity to increase health literacy, self-efficacy, and treatment adherence among populations [3-9]. However, these public tools have also opened the door to unprecedented social and health risks [10,11]. Although these platforms have demonstrated usefulness for health promotion [7,12], recent studies have suggested that false or misleading health information may spread more easily than scientific knowledge through social media [13,14]. Therefore, it is necessary to understand how health misinformation spreads and how it can affect decision-making and health behaviors [15].

Although the term “health misinformation” is increasingly present in our societies, its definition is becoming increasingly elusive owing to the inherent dynamism of the social media ecosystem and the broad range of health topics [16]. Using a broad term that can include the wide variety of definitions in scientific literature, we here define health misinformation as a health-related claim that is based on anecdotal evidence, false, or misleading owing to the lack of existing scientific knowledge [1]. This general definition would consider, on the one hand, information that is false but not created with the intention of causing harm (ie, misinformation) and, on the other, information that is false or based on reality but deliberately created to harm a particular person, social group, institution, or country (ie, disinformation and malinformation).

The fundamental role of health misinformation on social media has been recently highlighted by the COVID-19 pandemic, as well as the need for quality and veracity of health messages in order to manage the present public health crisis and the subsequent infodemic. In fact, at present, the propagation of health misinformation through social media has become a major public health concern [17]. The lack of control over health information on social media is used as evidence for the current demand to regulate the quality and public availability of online information [18]. In fact, although today there is broad agreement among health professionals and policy makers on the need to control health misinformation, there is still little evidence about the effects that the dissemination of false or misleading health messages through social media could have on public health in the near future. Although recent studies are exploring innovative ways to effectively combat health misinformation online [19-22], additional research is needed to characterize and capture this complex social phenomenon [23].

More specifically, four knowledge gaps have been detected from the field of public health [1]. First, we have to identify the dominant health misinformation trends and specifically assess their prevalence on different social platforms. Second, we need to understand the interactive mechanisms and factors that make it possible to progressively spread health misinformation through social media (eg, vaccination myths, miracle diets, alternative treatments based on anecdotal evidence, and misleading advertisements on health products). Factors, such as the sources of misinformation, structure and dynamics of online communities, idiosyncrasies of social media channels, motivation and profile of people seeking health information, content and framing of health messages, and context in which misinformation is shared, are critical to understanding the dynamics of health misinformation through these platforms. For instance, although the role of social bots in spreading misinformation through social media platforms during political campaigns and election periods is widely recognized, health debates on social media are also affected by social bots [24]. At present, social bots are used to promote certain products in order to increase company profits, as well as to benefit certain ideological positions or contradict health evidence (eg, in the case of vaccines) [25]. Third, a key challenge in epidemiology and public health research is to determine not only the effective impact of these tools in the dissemination of health misinformation but also their impact on the development and reproduction of unhealthy or dangerous behaviors. Finally, regarding health interventions, we need to know which strategies are the best in fighting and reducing the negative impact of health misinformation without reducing the inherent communicative potential to propagate health information with these same tools.

In line with the abovementioned gaps, a recent report represents one of the first steps forward in the comparative study of health misinformation on social media [16]. Through a systematic review of the literature, this study offers a general characterization of the main topics, areas of research, methods, and techniques used for the study of health misinformation. However, despite the commendable effort made to compose a comprehensible image of this highly complex phenomenon, the lack of objective indicators that make it possible to measure the problem of health misinformation is still evident today.

Taking into account this wide set of considerations, this systematic review aimed to specifically address the knowledge gap. In order to guide future studies in this field of knowledge, our objective was to identify and compare the prevalence of health misinformation topics on social media platforms, with specific attention paid to the methodological quality of the studies and the diverse analytical techniques that are being implemented to address this public health concern.

Methods

Guideline

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [26].

Inclusion Criteria

Studies were included if (1) the objectives were to address the study of health misinformation on social media, search systematically for health misinformation, and explicitly discuss the impact, consequences, or purposes of misinformation; (2) the results were based on empirical results and the study used quantitative, qualitative, and computational methods; and (3) the research was specifically focused on social media platforms (eg, Twitter, Facebook, Instagram, Flickr, Sina Weibo, VK, YouTube, Reddit, Myspace, Pinterest, and WhatsApp). For comparability, we included studies written in English that were published after 2000 until March 2019.

Exclusion Criteria

Articles were excluded if they addressed health information quality in general or if they partially mentioned the existence of health misinformation without providing empirical findings. We did not include studies that dealt with content posted on other social media platforms. During the screening process, papers with a lack of methodological quality were also excluded.

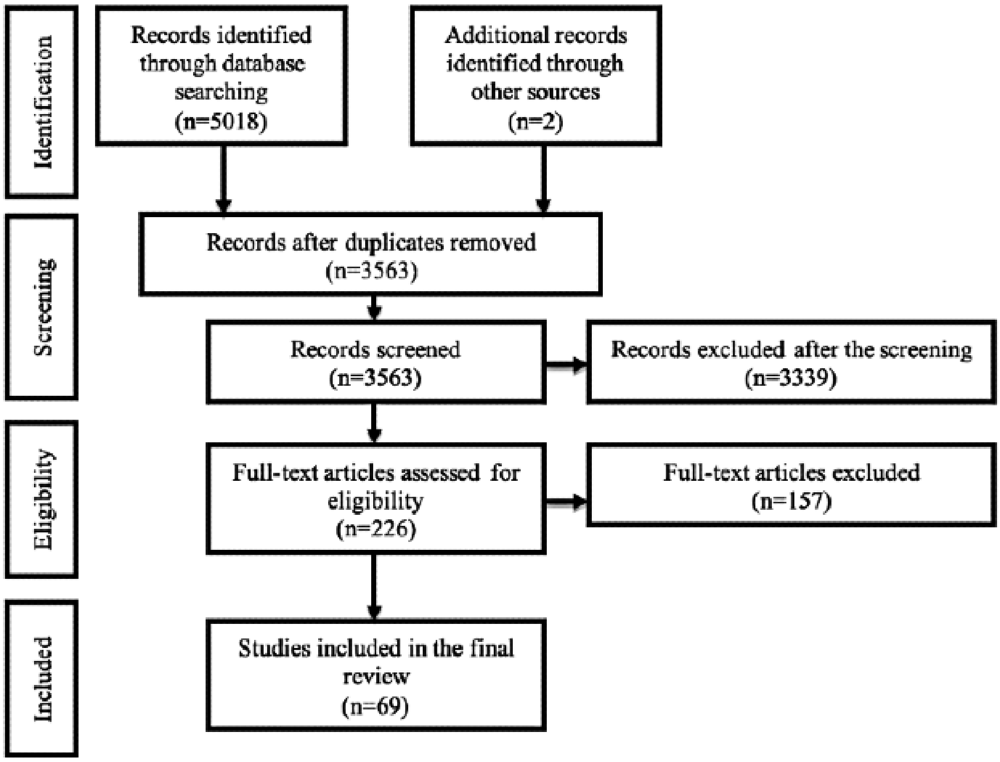
Search Strategy

We searched MEDLINE and PREMEDLINE in March 2019 using the PubMed search engine. Based on previous findings [16], the query searched for MeSH terms and keywords (in the entire body of the manuscript) related to the following three basic analytical dimensions that articulated our research objective: (1) social media, (2) health, and (3) misinformation. The MeSH terms were social media AND health (ie, this term included health behaviors) AND (misinformation OR information seeking behavior OR communication OR health knowledge, attitudes, practice). Based on the results obtained through this initial search, we added some keywords that (having been extracted from the articles that met the inclusion criteria) were specifically focused on the issue of health misinformation on social media. The search using MeSH terms was supplemented with the following keywords: social media (eg, "Twitter" OR "Facebook" OR "Instagram" OR "Flickr" OR "Sina Weibo" OR "YouTube" OR "Pinterest") AND health AND misinformation (eg, "inaccurate information" OR "poor quality information" OR "misleading information" OR "seeking information" OR "rumor" OR "gossip" OR "hoax" OR "urban legend" OR "myth" OR "fallacy" OR "conspiracy theory"). This initial search retrieved 1693 records. Additionally, this search strategy was adapted for its use in Scopus (3969 records) and Web of Science (1541 records). A full description of the search terms can be found in Multimedia Appendix 1.

Study Selection

In total, we collected 5018 research articles. After removing duplicates, we screened 3563 articles and retrieved 226 potentially eligible articles. In the next stage, we independently carried out a full-text selection process for inclusion (k=0.89). Discrepancies were shared and resolved by mutual agreement. Finally, a total of 69 articles were included in this systematic review (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart.



Data Extraction

In the first phase, the data were extracted by VSL and then checked by VSL and JAG. In order to evaluate the quality of the selected studies and given the wide variety of methodologies and approaches found in the articles, we composed an extraction form based on previous work [27-29]. Each extraction form contained 62 items, most of which were closed questions that could be answered using predefined forms (yes/good, no/poor, partially/fair, etc). Following this coding scheme, we extracted the following four different fields of information: (1) descriptive information (27 items), (2) search strategy evaluation (eight items), (3) information evaluation (six items), and (4) the quality and rigor of methodology and reporting (15 items) for either quantitative or qualitative studies (Multimedia Appendix 1). Questions in field 2, which have been used in previous studies [27], assessed the quality of information provided to demonstrate how well reported, systematic, and comprehensive the search strategy was (S score). The items in field 3 measured how rigorous the evaluation was (E score) for health-related misinformation [27]. Field 4 contained items designed for the general evaluation of quality in the research process, whether quantitative [28] or qualitative [29]. This Q-score approach takes into account general aspects of the research and reporting, such as the study, methodology, and quality of the discussion. For each of the information fields, we calculated the raw score as the sum of each of the items by equating “yes” or “good” as 1 point, “fair” as 0.5 points, and “no” or “poor” as 0 points (Multimedia Appendix 2). The purpose of these questions is to guarantee the quality of the selected studies.

Furthermore, in order to be able to compare the methods used in the selected studies, the studies were classified into several categories. The studies classified as “content/text analysis” used methods related to textual and content analysis, emphasizing the word/topic frequency, linguistic inquiry word count, n-grams, etc. The second category “evaluating content” grouped together studies whose methods were focused on the evaluation of content and information. In general, these studies analyzed different dimensions of the information published on social media. The third category “evaluating quality” included studies that analyzed the quality of the information offered in a global way. This category considered other dimensions in addition to content, such as readability, accuracy, usefulness, and sources of information. The fourth category “sentiment analysis” included studies whose methods were focused on sentiment analysis techniques (ie, methods measuring the reactions and the general tone of the conversation on social media). Finally, the “social network analysis” category included those studies whose methods were based on social network analysis techniques. These studies focused on measuring how misinformation spreads on social media, the relationship between the quality of information and its popularity on these social platforms, the relationship between users and opinions, echochambers effects, and opinion formation.

Of the 226 studies available for full-text review, 157 were excluded for various reasons, including research topics that were not focused on health misinformation (n=133). We also excluded articles whose research was based on websites rather than social media platforms (n=16), studies that did not assess the quality of health information (n=6) or evaluated institutional communication (n=5), nonempirical studies (n=2), and research protocols (n=1). In addition, two papers were excluded because of a lack of quality requirements (Q score <50%). Finally, the protocol of this review was registered at the International Prospective Register of Systematic Reviews (PROSPERO CRD42019136694).

Results

Prevalence of Health Misinformation

Ultimately, 69 studies were identified as eligible, and they covered a wide range of health topics and social media platforms, with the most common data source being Twitter (29/69, 43%), followed by YouTube (25/69, 37%) and Facebook (6/69, 9%). The less common sources were Instagram, MySpace, Pinterest, Tumblr, WhatsApp, and VK or a combination of these. Overall, 90% (61/69) of the studies were published in health science journals, and only 7% (5/69) of the studies were published in communication journals. The vast majority of articles analyzed posts written exclusively in one language (63/69, 91%). Only a small percentage assessed posts in more than one language (6/69, 10%).

Table 1 classifies the studies by topic and social media platform [30-97]. It also includes the prevalence of health misinformation posts. The topics were articulated around the following six principal categories: vaccines (22/69, 32%), drugs or smoking issues (16/69, 22%), noncommunicable diseases (13/69, 19%), pandemics (7/69, 10%), eating disorders (6/69, 9%), and medical treatments (5/69, 7%). The quality assessment results for the S score, E score, and Q score are reported in Multimedia Appendix 3.

Figure 2 shows the prevalence of health misinformation grouped by different topics and social media typology. Studies are ordered according to the percentage of health misinformation posts found in the studies selected. These works were also classified according to the type of social media under study. In this way, papers focused on Twitter, Tumblr, or Myspace were categorized as “microblogging.” Additionally, papers focused on YouTube, Pinterest, or Instagram were classified within “media sharing” platforms. Moreover, papers focused on Facebook, VK, or WhatsApp were included within the group of “social network” platforms. While all topics were present on all the different social media platforms, we found some differences in their prevalence. On one hand, vaccines, drugs, and pandemics were more prevalent topics on microblogging platforms (ie, Twitter or MySpace). On the other hand, on media sharing platforms (ie, YouTube, Instagram, or Pinterest) and social network platforms (ie, Facebook, VK, or WhatsApp), noncommunicable diseases and treatments were the most prevalent topics. More specifically, Twitter was the most used source for work on vaccines (10/69), drugs or smoking products (10/69), pandemics (4/69), and eating disorders (3/69). For studies on noncommunicable diseases (9/69) or treatments (5/69), YouTube was the most used social media platform.

Table 1. Summary of the prevalence of misinformation by topic and social media platform.

Authors	Year	Topic	Social media platform	Prevalence of health misinformation posts
Abukaraky et al [30]	2018	Treatments	YouTube	30%
Ahmed et al [31]	2019	Pandemics	Twitter	N/A[a]
Al Khaja et al [32]	2018	Drugs	WhatsApp	27%
Allem et al [33]	2017	Drugs	Twitter	59%
Allem et al [34]	2017	Drugs	Twitter	N/A
Arseniev-Koehler et al [35]	2016	EDs[b]	Twitter	36%
Basch et al [36]	2017	Vaccines	YouTube	65%
Becker et al [37]	2016	Vaccines	Twitter	1%
Biggs et al [38]	2013	NCDs[c]	YouTube	39%
Blankenship et al [39]	2018	Vaccines	Twitter	24%
Bora et al [40]	2018	Pandemics	YouTube	23%
Branley et al [41]	2017	EDs	Twitter and Tumblr	25%
Briones et al [42]	2012	Vaccines	YouTube	51%
Broniatowski et al [23]	2018	Vaccines	Twitter	35%
Buchanan et al [43]	2014	Vaccines	Facebook	43%
Butler et al [44]	2013	Treatments	YouTube	N/A
Cavazos-Rehg et al [45]	2018	Drugs	Twitter	75%
Chary et al [46]	2017	Drugs	Twitter	0%
Chew et al [47]	2010	Pandemics	Twitter	4%
Covolo et al [48]	2017	Vaccines	YouTube	23%
Dunn et al [49]	2015	Vaccines	Twitter	25%
Dunn et al [50]	2017	Vaccines	Twitter	N/A
Ekram et al [51]	2018	Vaccines	YouTube	57%
Erdem et al [52]	2018	Treatments	YouTube	0%
Faasse et al [53]	2016	Vaccines	Facebook	N/A
Fullwood et al [54]	2016	Drugs	YouTube	34%
Garg et al [55]	2015	Vaccines	YouTube	11%
Gimenez-Perez et al [56]	2018	NCDs	YouTube	50%
Goobie et al [57]	2019	NCDs	YouTube	N/A
Guidry et al [58]	2017	Pandemics	Twitter and Instagram	N/A
Guidry et al [59]	2016	Drugs	Pinterest	97%
Guidry et al [60]	2015	Vaccines	Pinterest	74%
Hanson et al [61]	2013	Drugs	Twitter	0%
Harris et al [62]	2018	EDs	Twitter	N/A
Haymes et al [63]	2016	NCDs	YouTube	47%
Helmi et al [64]	2018	NCDs	Different sources	N/A
Kang et al [65]	2017	Vaccines	Twitter	42%
Katsuki et al [66]	2015	Drugs	Twitter	6%
Keelan et al [67]	2010	Vaccines	MySpace	43%
Keim-Malpass et al [68]	2017	Vaccines	Twitter	43%
Kim et al [69]	2017	NCDs	YouTube	22%

Authors	Year	Topic	Social media platform	Prevalence of health misinformation posts
Krauss et al [70]	2017	Drugs	Twitter	50%
Krauss et al [71]	2015	Drugs	Twitter	87%
Kumar et al [72]	2014	NCDs	YouTube	33%
Laestadius et al [73]	2016	Drugs	Instagram	N/A
Leong et al [74]	2018	NCDs	YouTube	33%
Lewis et al [75]	2015	Treatments	YouTube	N/A
Loeb et al [76]	2018	NCDs	YouTube	77%
Love et al [77]	2013	Vaccines	Twitter	13%
Martinez et al [78]	2018	Drugs	Twitter	67%
Massey et al [79]	2016	Vaccines	Twitter	25%
McNeil et al [80]	2012	NCDs	Twitter	41%
Menon et al [81]	2017	Treatments	YouTube	2%
Merianos et al [82]	2016	Drugs	YouTube	65%
Meylakhs et al [83]	2014	NCDs	VK	N/A
Morin et al [84]	2018	Pandemics	Twitter	N/A
Mueller et al [85]	2019	NCDs	YouTube	66%
Porat et al [86]	2019	Pandemics	Twitter	0%
Radzikowski et al [87]	2016	Vaccines	Twitter	N/A
Schmidt et al [88]	2018	Vaccines	Facebook	4%
Seltzer et al [89]	2017	Pandemics	Instagram	60%
Seymour et al [90]	2015	NCDs	Facebook	N/A
Syed-Abdul et al [91]	2013	EDs	YouTube	29%
Teufel et al [92]	2013	EDs	Facebook	22%
Tiggermann et al [93]	2018	EDs	Twitter	29%
Tuells et al [94]	2015	Vaccines	YouTube	12%
van der Tempel et al [95]	2016	Drugs	Twitter	N/A
Waszak et al [96]	2018	NCDs	Facebook	40%
Yang et al [97]	2018	Drugs	YouTube	98%

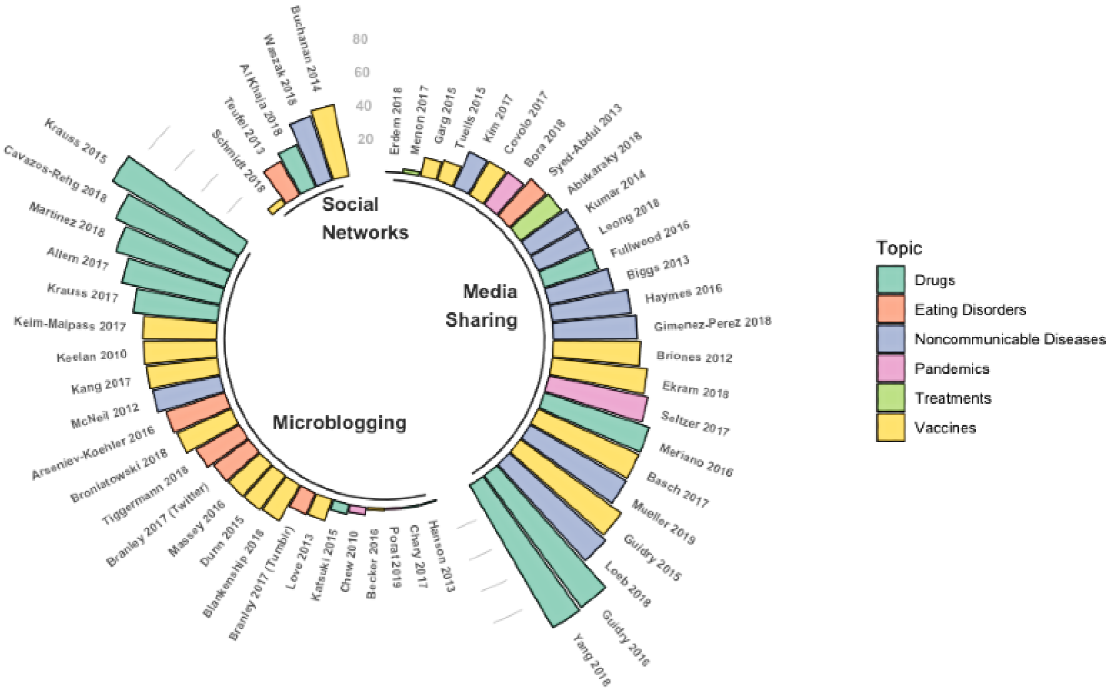
[a] N/A: not applicable.

[b] EDs: eating disorders.

[c] NCDs: noncommunicable diseases.

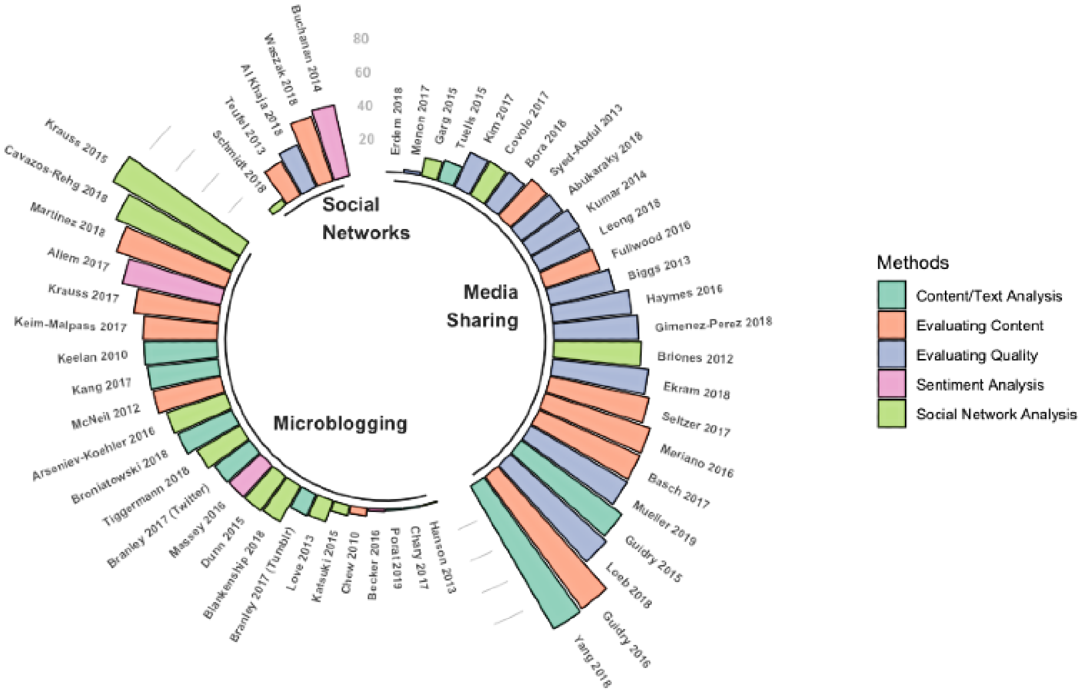
Overall, health misinformation was most prevalent in studies related to smoking products, such as hookah and water pipes [33,59,71], e-cigarettes, and drugs, such as opioids and marijuana [45,70,97]. Health misinformation about vaccines was also very common. However, studies reported different levels of health misinformation depending on the type of vaccine studied, with the human papilloma virus (HPV) vaccine being the most affected [67,68]. Health misinformation related to diets or pro–eating disorder arguments were moderate in comparison to the aforementioned topics [35,93]. Studies focused on diseases (ie, noncommunicable diseases and pandemics) also reported moderate misinformation rates [56,85], especially in the case of cancer [76,96]. Finally, the lowest levels of health misinformation were observed in studies evaluating the presence of health misinformation regarding medical treatments. Although first-aid information on burns or information on dental implants was limited in quantity and quality, the prevalence of misinformation for these topics was low. Surgical treatment misinformation was the least prevalent. This was due to the fact that the content related to surgical treatments mainly came from official accounts, which made the online information complete and reliable.

Figure 2. Prevalence of health misinformation grouped by different topics and social media type.



Regarding the methods used in the different studies, there were some differences between the diverse social media platforms. We classified the studies based on the methods applied into the following five categories: social network analysis (19/69), evaluating content (18/69), evaluating quality (16/69), content/text analysis (12/69), and sentiment analysis (4/69). Figure 3 shows the different methods applied in the studies classified by the type of social media platform and ordered by the percentage of misinformation posts. Among platforms, such as YouTube and Instagram, methods focused on the evaluation of health information quality and content were common, representing 22% (15/69) and 12% (8/69), respectively. On microblogging platforms, such as Twitter and Tumblr, social network analysis was the method most used by 19% (13/69) of the studies. Finally, on social media platforms, such as Facebook, VK, and WhatsApp, studies whose methods were related to social network analysis represented 3% (2/69) of the included studies and those focused on the evaluation of content represented 4% (3/69) of the included studies.

Figure 3. Prevalence of health misinformation grouped by methods and social media type.



Misinformation Topics and Methods

Vaccines

Overall, 32% (22/69) of the studies focused on vaccines or vaccination decision-making–related topics. Additionally, 14% (10/69) of the selected articles focused on social media discussion regarding the potential side effects of vaccination [23,36,48,53,55,60,65,77,87,88], 12% (8/69) were centered on the debate around the HPV vaccine [42,49-51,67,68,79,94], and 3% (2/69) were centered on the antivaccine movement [39,43]. According to social media platforms, 9% (6/69) of the studies were focused on the debate and narratives about vaccines in general on Twitter, and 6% (4/69) specifically analyzed the HPV debate on this platform. Papers focused on YouTube also followed a similar trend, and they were centered on the HPV debate and on the public discussion on vaccine side effects and risks for specific population groups (eg, autism in children). Regarding Facebook, all studies were particularly focused on vaccination decision-making.

Most authors studied differences in language use, the effect of a heterogeneous community structure in the propagation of health misinformation, and the role played by fake profiles or bots in the spread of poor quality, doubtful, or ambiguous health content. In line with these concerns, authors pointed out the need to further study the circumstances surrounding those who adopt these arguments [49], and whether alternative strategies to education could improve the fight against antivaccine content [51]. Authors also recommended paying close attention to social media as these tools are assumed to play a fundamental role in the propagation of misinformation. For instance, the role played by the echochamber or the heterogeneous community structure on Twitter has been shown to skew the information to which users are exposed in relation to HPV vaccines [49]. In this sense, it is widely acknowledged that health professionals should pay more attention to antivaccine arguments on social media, so that they can better respond to patients' concerns [36,43,65,77]. Furthermore, governmental organizations could also use social media platforms to reach a greater number of people [39,55].

Drugs and Smoking

Several studies (16/69, 22%) covered misuse and misinformation about e-cigarettes, marijuana, opioid consumption, and prescription drug abuse. Studies covering the promotion of e-cigarette use and other forms of smoking, such as hookah (ie, water pipes or narghiles) represented 7% (5/69) of the articles analyzed. The rest (16%, 11/69) were focused on the analysis of drug misinformation.

According to topic, regarding drug and opioid use, studies investigated the dissemination of misinformation through social media platforms [32,45,46,70,97], the consumption of misinformation related to these products, drug abuse, and the sale of online medical products [61,66]. These studies highlighted the risk, especially for young people, caused by the high rate of misinformation related to the dissemination of drug practice and misuse (predominantly marijuana and opioids) [45]. In addition, social media platforms were identified as a potential source of illegal promotion of the sale of controlled substances directly to consumers [66]. Most drug-related messages on social media were potentially misleading or false claims that lacked credible evidence to support them [32]. Other studies pointed to social media as a potential source of information that illegally promotes the sale of controlled prescription drugs directly to consumers [66]. In the case of cannabinoids, there was often content that described, encouraged, promoted [54], or even normalized the consumption of illicit substances [70].

Unlike drug studies, most of the papers analyzed how e-cigarettes and hookah [33,34,59,71,73,78,82,95] are portrayed on social media and/or the role of bots in promoting e-cigarettes. Regarding e-cigarettes, studies pointed out the high prevalence of misinformation denying health damage [95]. In this sense, it is worth noting the importance of sources of misinformation. While in the case of vaccines, the source of health misinformation was mainly individuals or groups of people with a particular interest (eg, antivaccine movement), social media was found to be frequently contaminated by misinformation from bots (ie, software applications that autonomously run tasks such as spreading positive discourse about e-cigarettes and other tobacco products) [78]. In fact, these fake accounts may influence the online conversation in favor of e-cigarettes given the scientific appearance of profiles [78]. Some of the claims found in this study denied the harmfulness of e-cigarettes. In line with these findings, other studies pointed to the high percentage of messages favoring e-cigarettes as an aid to quitting smoking [95].

We found that 10% (7/69) of the studies used methods focused on evaluating the content of the posts. These studies aimed to explore the misperceptions of drug abuse or alternative forms of tobacco consumption. Along these lines, another study (1/69, 1%) focused on evaluating the quality of content. The authors evaluated the truthfulness of claims about drugs. In particular, we found that 7% (5/69) of the studies used social network analysis techniques. These studies analyzed the popularity of messages based on whether they promoted illegal access to drugs online and the interaction of users with this content. Other studies (3/69, 3%) used content analysis techniques. These studies evaluated the prevalence of misinformation on platforms and geographically, as a kind of "toxicosurveillance" system [34,46].

Noncommunicable Diseases

A relevant proportion (13/69, 19%) of studies assessed noncommunicable diseases, such as cancer, diabetes, and epilepsy. Most of the studies focused on the objective evaluation of information quality on YouTube [38,56,57,69,72,74,76,80,85]. Overall, 13% (9/69) of these studies used methods to assess the quality of the information. The authors analyzed the usefulness and accuracy of the information. Moreover, 4% (3/69) of the studies used methods related to content assessment. The main objective of these studies was to analyze which are the most common misinformation topics. Furthermore, 3% (2/69) used social network analysis, and the main objective of the analysis was to study the information dissemination patterns or the social spread of scientifically inaccurate health information.

Some studies evaluated the potential of this platform as a source of information specially for health students or self-directed education among the general public. Unfortunately, the general tone of research findings was that YouTube is not an advisable source for health professionals or health information seekers. Regarding diabetes, the probability of finding misleading videos was high [56]. Misleading videos promoted cures for diabetes, negated scientific arguments, or provided treatments with no scientific basis. Furthermore, misleading videos related to diabetes were found to be more popular than those with evidence-based health information [74], which increased the probability of consuming low-quality health content. The same misinformation pattern was detected for other chronic diseases such as hypertension [72], prostate cancer [76], and epilepsy [80].

Pandemics and Communicable Diseases

Results indicated that 10% (7/69) of the studies covered misinformation related to pandemics and communicable diseases such as H1N1 [31,47], Zika [40,89], Ebola [58,84], and diphtheria [86]. All these studies analyzed how online platforms were used by both health information seekers and health and governmental authorities during the pandemic period.

We found that 14% (10/69) of the studies on this topic evaluated the quality of the information. To achieve this, most of the studies used external instruments such as DISCERN and AAD7 Self-Care Behaviors. Overall, 9% (6/69) of the papers evaluated the content of the information. These studies were focused on analysis of the issues of misinformation. Another 4% (3/69) used social media analysis to observe the propagation of misinformation. Finally, 3% (2/69) used textual analysis as the main method. These studies focused on the study of the prevalence of health misinformation.

These studies identified social media as a public forum for free discussion and indicated that this freedom might lead to rumors on anecdotal evidence and misunderstandings regarding pandemics. Consequently, although social media was described as a forum for sharing health-related knowledge, these tools are also recognized by researchers and health professionals as a source of misinformation that needs to be controlled by health experts [83,84]. Therefore, while social media serves as a place where people commonly share their experiences and concerns, these platforms can be potentially used by health professionals to fight against false beliefs on communicable diseases (eg, as it is happening today during the COVID-19 pandemic). Accordingly, social media platforms have been found to be powerful tools for health promotion among governmental institutions and health-related workers, and new instruments that, for instance, are being used to increase health surveillance and intervention against false beliefs and misinformation [31,89]. In fact, different authors agreed that governmental/health institutions should increase their presence on social media platforms during pandemic crises [47,58,84,86].

Diet/Eating Disorders

Studies focusing on diet and eating disorders represented 9% (6/69) of the included studies. This set of studies identified pro–eating disorder groups and discourses within social media [35], and how pro–eating disorder information was shared and spread on these platforms [91]. Anorexia was the most studied eating disorder along with bulimia. Furthermore, discourses promoting fitness or recovery after an eating disorder were often compared with those issued by pro–eating disorder groups [41,62,92,93]. In general, the authors agreed on the relevance of pro–eating disorder online groups, the mutual support among members, and the way they reinforce their opinions and health behaviors [35].

Overall, 4% (3/69) of the studies used social network analysis techniques. The authors focused on analyzing the existing connections between individuals in the pro–eating disorder community and their engagement, or comparing the cohesion of these communities with other communities, such as the fitness community, that promote healthier habits. Moreover, 3% (2/69) of the studies evaluated the quality of the content and particularly focused on informative analysis of the videos, that is, the content was classified as informative when it described the health consequences of anorexia or proana if, on the contrary, anorexia was presented as a fashion or a source of beauty. Furthermore, only one study used content analysis techniques. The authors classified the posts according to the following categories: proana, antiana, and prorecovery. Pro–eating disorder pages tended to identify themselves with body-associated pictures owing to the importance they attributed to motivational aspects of pro–eating disorder communities [92]. The pro–eating disorder claims contained practices about weight loss, wanting a certain body type or characteristic of a body part, eating disorders, binge eating, and purging [62]. Pro–eating disorder conversations also had a high content of social support in the form of tips and tricks (eg, “Crunch on some ice chips if you are feeling a hunger craving. This will help you feel as if you are eating something substantial” and “How do you all feel about laxatives?”) [92].

Regarding eating disorders on social media, paying attention to community structure is important according to authors. Although it is widely acknowledged that communities can be positive by providing social support, such as recovery and well-being, certain groups on social media may also reaffirm the pro–eating disorder identity [35]. In fact, polarized pro/anti–eating disorder communities can become closed echochambers where community members are selectively exposed to the content they are looking for and therefore only hear the arguments they want to hear. In this case, the echochamber effect might explain why information campaigns are limited in scope and often encourage polarization of opinion, and can even reinforce existing divides in pro–eating disorder opinions [88].

Treatments and Medical Interventions

Finally, we found that 7% (5/69) of the studies assessed the quality of health information regarding different medical treatments or therapies recommended through social media [63,81]. According to method, 6% (4/69) of the studies evaluated the quality of information related to the proposed treatments and therapies. In this sense, the fundamental goal of these studies was aimed at assessing the quality and accuracy of the information.

As in the case of noncommunicable diseases, professionals scanned social networks, especially YouTube, and evaluated the quality of online health content as an adequate instrument for self-care or for health student training. There were specific cases where information was particularly limited in quality and quantity, such as dental implants and first-aid information on burns [30,44]. However, most surgical treatments or tools were found to have a sufficient level of quality information on YouTube [52,81]. In relation to this topic, it is worth pointing out the source of the misinformation. In this particular case, most of the posts were published by private companies. They used the platforms to promote their medical products. Therefore, the amount of misinformation was considerably low compared with other topics, such as eating disorders and vaccines, that are closely linked to the general public. In general, the videos were accurate, were well presented, and framed treatments in a useful way for both health workers and health information seekers.

A full description of the objectives and main conclusions of the reviewed articles is presented in Multimedia Appendix 4.

Discussion

Main Findings

This work represents, to our knowledge, the first effort aimed at finding objective and comparable measures to quantify the extent of health misinformation in the social media ecosystem. Our study offers an initial characterization of dominant health misinformation topics and specifically assesses their prevalence on different social platforms. Therefore, our systematic review provides new insights on the following unanswered question that has been recurrently highlighted in studies of health misinformation on social media: How prevalent is health misinformation for different topics on different social platform types (ie, microblogging, media sharing, and social networks)?

We found that health misinformation on social media is generally linked to the following six topical domains: (1) vaccines, (2) diets and eating disorders, (3) drugs and new tobacco products, (4) pandemics and communicable diseases, (5) noncommunicable diseases, and (6) medical treatments and health interventions.

With regard to vaccines, we found some interesting results throughout the different studies. Although antivaccine ideas have been traditionally linked to emotional discourse against the rationality of the scientific and expert community, we curiously observed that in certain online discussions, antivaccine groups tend to incorporate scientific language in their own discourse with logically structured statements and/or with less usage of emotional expressions [53]. Thus, the assimilation of the scientific presentation and its combination with anecdotal evidence can rapidly spread along these platforms through a progressive increment of visits and “likes” that can make antivaccine arguments particularly convincing for health information seekers [53,55]. Furthermore, we found that the complex and heterogeneous community structure of these online groups must be taken into account. For instance, those more exposed to antivaccine information tend to spread more negative concerns about vaccines (ie, misinformation or opinions related to vaccine hesitancy) than users exposed to positive or neutral opinions [49]. Therefore, negative/positive opinions are reinforced through the network structure of particular social media platforms. Moreover, fake profiles tend to amplify the debate and discussion, thereby undermining the possible public consensus on the effectiveness and safety of vaccines, especially in the case of HPV; measles, mumps, and rubella (MMR); and influenza [23].

As observed in our review, health topics were omnipresent over all social media platforms included in our study; however, the health misinformation prevalence for each topic varied depending on platform characteristics. Therefore, the potential effect on population health was ambivalent, that is, we found both positive and negative effects depending on the topic and on the group of health information seekers. For instance, content related to eating disorders was frequently hidden or not so evident to the general public, since pro–eating disorder communities use their own codes to reach specific audiences (eg, younger groups) [98]. To provide a simple example, it is worth mentioning the usage of nicknames, such as proana for proanorexia and promia for probulimia, as a way to reach people with these health conditions and make it easier for people to talk openly about their eating disorders. More positively, these tools have been useful in prevention campaigns during health crises. For example, during the H1N1, Ebola, and Zika pandemics, and, even more recently, with the ongoing COVID-19 pandemic, platforms, such as Twitter, have been valuable instruments for spreading evidence-based health knowledge, expert recommendations, and educative content aimed at avoiding the propagation of rumors, risk behaviors, and diseases [31,89].

Throughout our review, we found different types of misinformation claims depending on the topic. Concerning vaccines, misinformation was often framed with a scientific appearance against scientific evidence [53]. Drug-related misinformation promoted the consumption and abuse of these substances [66]. However, these statements lacked scientific evidence to support them [32]. As with vaccines, false accounts that influenced the online conversation did so with a scientific appearance in favor of e-cigarettes [82]. In this sense, most accounts tended to promote the use and abuse of these items. With beauty as the final goal, misinformation about eating disorders promoted changes in the eating habits of social media users [91]. Furthermore, we found that social media facilitated the development of pro–eating disorder online communities [35]. In general, the results indicated that this type of content promoted unhealthy practices while normalizing eating disorders. In contrast, epidemic/pandemic-related misinformation was not directly malicious. Misinformation on this topic involved rumors, misunderstandings, and doubts arising from a lack of scientific knowledge [31]. The statements were within the framework of the health emergency arising from the pandemic. In line with these findings, we noted findings related to noncommunicable diseases. Messages that focused on this topic promoted cures for chronic diseases or for conditions with no cure through fallacies or urban legends [85].

In this study, we focused on analysis of the results obtained and the conclusions of the authors. Some of our findings are in line with those obtained in recent works [16]. The reviewed studies indicate, on one hand, the difficulty in characterizing and evaluating the quality of health information on social media [1] and, on the other, the conceptual fuzziness that can result from the convergence of multiple disciplines trying to apprehend the multidisciplinary and complex phenomenon of health misinformation on social media. This research field is being studied by health and social scientists [70,73], as well as by researchers from the fields of computer science, mathematics, sociophysics, etc [99,100]. Therefore, we must understand that the inherent multidisciplinary and methodological diversity of studies and the highly dynamic world of social media are a perfect match for making it more difficult to identify comprehensive and transversal solutions to the problem of health misinformation. In fact, as we have found, misinformation on vaccines, drugs, and new smoking products is more prevalent on media-sharing platforms (eg, YouTube) and microblogging applications (eg, Twitter), while misinformation on noncommunicable diseases is particularly prevalent on media sharing platforms where users can widely describe disease symptoms, medical treatments, and therapies [76,85]. Platforms, such as YouTube, owing to their characteristics, allow more space for users to share this type of information, while the natural dynamism of Twitter makes it an ideal medium for discussion among online communities with different political or ideological orientations (eg, pro/antivaccination communities).

Finally, we should mention that the current results are limited to the availability and quality of social media data. Although the digitalization of social life offers researchers an unprecedented amount of health and social information that can be used to understand human behaviors and health outcomes, accessing this online data is becoming increasingly difficult, and some measures have to be taken to mitigate bias [40,43,67,79]. Over the last few years, new concerns around privacy have emerged and led governments to tighten regulations around data access and storage [101,102]. Consequently, in response to these new directives, as well as scandals involving data sharing and data breaches such as the Cambridge Analytica case, social media companies are developing new controls and barriers to data in their platforms. This is why free access to application programming interfaces (APIs) is becoming increasingly difficult and the range of social data accessible via APIs is gradually decreasing. These difficulties in accessing data are also determining which platforms are most frequently used by researchers, which are not used, and which will be used in the near future.

Limitations and Strengths

The present study has some limitations. First, the conceptual definition of health misinformation is one limitation. In any case, taking into account that we were facing a new field of study, we considered a broad definition in order to be more inclusive and operative in the selection of studies. Therefore, we included as many papers as possible for the review in order to perform an analysis of the largest number of possible topics. Second, from a methodological perspective, our findings are limited to research published in English language journals and do not cover all the social media platforms that exist. Besides, we discovered some technical limitations when conducting this systematic review. Owing to the newness of this research topic, our study revealed difficulties in comparing different research studies characterized by specific theoretical approaches, working definitions, methodologies, data collection processes, and analytical techniques. Some studies selected involved observational designs (using survey methods and textual analysis), whereas others were based on the application of automatic or semiautomatic computational procedures with the aim of classifying and analyzing health misinformation on social media. Finally, taking into account the particular features of each type of social media (ie, microblogging service, video sharing service, or social network) and the progressive barriers in accessing social media data, we need to consider the information and selection bias when studying health misinformation on these platforms. According to these biases, we should ponder which users are behind these tools and how we can extrapolate specific findings (ie, applied to certain groups and social media platforms) to a broader social context.

Despite the limitations described above, it is necessary to mention the strengths of our work. First, we believe that this study represents one of the first steps in advancing research involving health misinformation on social media. Unlike previous work, we offer some measures that can serve as guidance and a comparative baseline for subsequent studies. In addition, our study highlights the need to redirect future research toward social media platforms, which, perhaps due to the difficulties of automatic data collection, are currently being neglected by researchers. Our study also highlights the need for both researchers and health professionals to explore the possibility of using these digital tools for health promotion and the need for them to progressively colonize the social media ecosystem with the ultimate goal of combating the waves of health misinformation that recurrently flood our societies.

Conclusion

Health misinformation was most common on Twitter and on issues related to smoking products and drugs. Although we should be aware of the difficulties inherent in the dynamic magnitude of online opinion flows, our systematic review offers a comprehensive comparative framework that identifies subsequent action areas in the study of health misinformation on social media. Despite the abovementioned limitations, our research presents some advances when compared with previous studies. Our study provides (1) an overview of the prevalence of health misinformation identified on different social media platforms; (2) a methodological characterization of studies focused on health misinformation; and (3) a comprehensive description of the current research lines and knowledge gaps in this research field.

According to the studies reviewed, the greatest challenge lies in the difficulty of characterizing and evaluating the quality of the information on social media. Knowing the prevalence of health misinformation and the methods used for its study, as well as the present knowledge gaps in this field will help us to guide future studies and, specifically, to develop evidence-based digital policy action plans aimed at combating this public health problem through different social media platforms.

Acknowledgments

We would like to acknowledge the support of the University Research Institute on Social Sciences (INDESS, University of Cadiz) and the Ramon & Cajal Program. JAG was subsidized by the Ramon & Cajal Program operated by the Ministry of Economy and Business (RYC-2016-19353) and the European Social Fund.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and results from the search query.
[DOCX File , 15 KB - jmir_v23i1e17187_app1.docx]

Multimedia Appendix 2
Data extraction sheet.
[DOCX File , 15 KB - jmir_v23i1e17187_app2.docx]

Multimedia Appendix 3
Summary of quality scores.
[DOCX File , 21 KB - jmir_v23i1e17187_app3.docx]

Multimedia Appendix 4
Summary table with objectives and conclusions about misinformation prevalence in social media.
[DOCX File , 41 KB - jmir_v23i1e17187_app4.docx]

References

1. Chou WS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
2. Xiong F, Liu Y. Opinion formation on social media: an empirical approach. Chaos 2014 Mar;24(1):013130. [doi: 10.1063/1.4866011] [Medline: 24697392]
3. Altevogt B, Stroud C, Hanson S, Hanfling D, Gostin LE. Guidance for Establishing Crisis Standards of Care for Use in Disaster Situations. Washington, DC: National Academies Press; 2009.
4. Funk S, Gilad E, Watkins C, Jansen VAA. The spread of awareness and its impact on epidemic outbreaks. Proc Natl Acad Sci U S A 2009 Apr 21;106(16):6872-6877 [FREE Full text] [doi: 10.1073/pnas.0810762106] [Medline: 19332788]
5. Funk S, Salathé M, Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 2010 Sep 06;7(50):1247-1256 [FREE Full text] [doi: 10.1098/rsif.2010.0142] [Medline: 20504800]
6. Kim SJ, Han JA, Lee T, Hwang T, Kwon K, Park KS, et al. Community-Based Risk Communication Survey: Risk Prevention Behaviors in Communities during the H1N1 crisis, 2010. Osong Public Health Res Perspect 2014 Feb;5(1):9-19 [FREE Full text] [doi: 10.1016/j.phrp.2013.12.001] [Medline: 24955307]
7. McGloin AF, Eslami S. Digital and social media opportunities for dietary behaviour change. Proc Nutr Soc 2015 May;74(2):139-148. [doi: 10.1017/S0029665114001505] [Medline: 25319345]
8. Scarcella C, Antonelli L, Orizio G, Rossmann C, Ziegler L, Meyer L, et al. Crisis Communication in the Area of Risk Management: The CriCoRM Project. J Public Health Res 2013 Sep 02;2(2):e20 [FREE Full text] [doi: 10.4081/jphr.2013.e20] [Medline: 25170491]
9. Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding Vaccine Refusal: Why We Need Social Media Now. Am J Prev Med 2016 Apr;50(4):550-552 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.002] [Medline: 26655067]
10. Cavallo DN, Chou WS, McQueen A, Ramirez A, Riley WT. Cancer Prevention and Control Interventions Using Social Media: User-Generated Approaches. Cancer Epidemiol Biomarkers Prev 2014 Aug 07;23(9):1953-1956. [doi: 10.1158/1055-9965.epi-14-0593]
11. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. PLoS One 2014 Oct 15;9(10):e110171 [FREE Full text] [doi: 10.1371/journal.pone.0110171] [Medline: 25333470]
12. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 01;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]

13. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
14. Levy J, Strombeck R. Health benefits and risks of the Internet. J Med Syst 2002 Dec;26(6):495-510. [doi: 10.1023/a:1020288508362] [Medline: 12385532]
15. Pagoto S, Waring ME, Xu R. A Call for a Public Health Agenda for Social Media Research. J Med Internet Res 2019 Dec 19;21(12):e16661 [FREE Full text] [doi: 10.2196/16661] [Medline: 31855185]
16. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
17. Venkatraman A, Mukhija D, Kumar N, Nagpal SJS. Zika virus misinformation on the internet. Travel Med Infect Dis 2016 Jul;14(4):421-422. [doi: 10.1016/j.tmaid.2016.05.018] [Medline: 27267799]
18. Freeman B, Kelly B, Baur L, Chapman K, Chapman S, Gill T, et al. Digital junk: food and beverage marketing on Facebook. Am J Public Health 2014 Dec;104(12):e56-e64. [doi: 10.2105/AJPH.2014.302167] [Medline: 25322294]
19. Bode L, Vraga EK. See Something, Say Something: Correction of Global Health Misinformation on Social Media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
20. Bode L, Vraga EK. In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. J Commun 2015 Jun 23;65(4):619-638. [doi: 10.1111/jcom.12166]
21. Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. Vaccine 2015 Jan 09;33(3):459-464. [doi: 10.1016/j.vaccine.2014.11.017] [Medline: 25499651]
22. Vraga EK, Bode L. I do not believe you: how providing a source corrects health misperceptions across social media platforms. Information, Communication & Society 2017 Apr 19;21(10):1337-1353. [doi: 10.1080/1369118x.2017.1313883]
23. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
24. Allem J, Ferrara E. Could Social Bots Pose a Threat to Public Health? Am J Public Health 2018 Aug;108(8):1005-1006. [doi: 10.2105/AJPH.2018.304512] [Medline: 29995482]
25. Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. Vaccine 2012 May 28;30(25):3727-3733. [doi: 10.1016/j.vaccine.2012.02.025] [Medline: 22365840]
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
28. Barratt H, Rojas-García A, Clarke K, Moore A, Whittington C, Stockton S, et al. Epidemiology of Mental Health Attendances at Emergency Departments: Systematic Review and Meta-Analysis. PLoS One 2016 Apr 27;11(4):e0154449 [FREE Full text] [doi: 10.1371/journal.pone.0154449] [Medline: 27120350]
29. Rojas-García A, Turner S, Pizzo E, Hudson E, Thomas J, Raine R. Impact and experiences of delayed discharge: A mixed-studies systematic review. Health Expect 2018 Feb 12;21(1):41-56 [FREE Full text] [doi: 10.1111/hex.12619] [Medline: 28898930]
30. Abukaraky A, Hamdan A, Ameera M, Nasief M, Hassona Y. Quality of YouTube TM videos on dental implants. Med Oral Patol Oral Cir Bucal 2018 Jul 01;23(4):e463-e468 [FREE Full text] [doi: 10.4317/medoral.22447] [Medline: 29924766]
31. Ahmed W, Bath PA, Sbaffi L, Demartini G. Novel insights into views towards H1N1 during the 2009 Pandemic: a thematic analysis of Twitter data. Health Info Libr J 2019 Mar 20;36(1):60-72. [doi: 10.1111/hir.12247] [Medline: 30663232]
32. Al Khaja KAJ, AlKhaja AK, Sequeira RP. Drug information, misinformation, and disinformation on social media: a content analysis study. J Public Health Policy 2018 Aug 24;39(3):343-357. [doi: 10.1057/s41271-018-0131-2] [Medline: 29795521]
33. Allem J, Ramanujam J, Lerman K, Chu K, Boley Cruz T, Unger JB. Identifying Sentiment of Hookah-Related Posts on Twitter. JMIR Public Health Surveill 2017 Oct 18;3(4):e74 [FREE Full text] [doi: 10.2196/publichealth.8133] [Medline: 29046267]
34. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
35. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. #Proana: Pro-Eating Disorder Socialization on Twitter. J Adolesc Health 2016 Jun;58(6):659-664. [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]
36. Basch CH, Zybert P, Reeves R, Basch CE. What do popular YouTube videos say about vaccines? Child Care Health Dev 2017 Jul;43(4):499-503. [doi: 10.1111/cch.12442] [Medline: 28105642]
37. Becker BFH, Larson HJ, Bonhoeffer J, van Mulligen EM, Kors JA, Sturkenboom MCJM. Evaluation of a multinational, multilingual vaccine debate on Twitter. Vaccine 2016 Dec 07;34(50):6166-6171. [doi: 10.1016/j.vaccine.2016.11.007] [Medline: 27840012]

38. Biggs TC, Bird JH, Harries PG, Salib RJ. YouTube as a source of information on rhinosinusitis: the good, the bad and the ugly. J Laryngol Otol 2013 Aug;127(8):749-754. [doi: 10.1017/S0022215113001473] [Medline: 23866821]
39. Blankenship EB, Goff ME, Yin J, Tse ZTH, Fu K, Liang H, et al. Sentiment, Contents, and Retweets: A Study of Two Vaccine-Related Twitter Datasets. Perm J 2018;22:17-138 [FREE Full text] [doi: 10.7812/TPP/17-138] [Medline: 29911966]
40. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health 2018 Sep 29;112(6):320-328 [FREE Full text] [doi: 10.1080/20477724.2018.1507784] [Medline: 30156974]
41. Branley DB, Covey J. Pro-ana versus Pro-recovery: A Content Analytic Comparison of Social Media Users' Communication about Eating Disorders on Twitter and Tumblr. Front Psychol 2017 Aug 11;8:1356 [FREE Full text] [doi: 10.3389/fpsyg.2017.01356] [Medline: 28848472]
42. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health Commun 2012 Jul;27(5):478-485. [doi: 10.1080/10410236.2011.610258] [Medline: 22029723]
43. Buchanan R, Beckett RD. Assessment of vaccination-related information for consumers available on Facebook. Health Info Libr J 2014 Sep;31(3):227-234 [FREE Full text] [doi: 10.1111/hir.12073] [Medline: 25041499]
44. Butler DP, Perry F, Shah Z, Leon-Villapalos J. The quality of video information on burn first aid available on YouTube. Burns 2013 Aug;39(5):856-859. [doi: 10.1016/j.burns.2012.10.017] [Medline: 23273651]
45. Cavazos-Rehg P, Zewdie K, Krauss M, Sowles S. "No High Like a Brownie High": A Content Analysis of Edible Marijuana Tweets. Am J Health Promot 2018 May;32(4):880-886. [doi: 10.1177/0890117116686574] [Medline: 29214836]
46. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. J Med Toxicol 2017 Dec 22;13(4):278-286 [FREE Full text] [doi: 10.1007/s13181-017-0625-5] [Medline: 28831738]
47. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
48. Covolo L, Ceretti E, Passeri C, Boletti M, Gelatti U. What arguments on vaccinations run through YouTube videos in Italy? A content analysis. Hum Vaccin Immunother 2017 Jul 03;13(7):1693-1699 [FREE Full text] [doi: 10.1080/21645515.2017.1306159] [Medline: 28362544]
49. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
50. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
51. Ekram S, Debiec KE, Pumper MA, Moreno MA. Content and Commentary: HPV Vaccine and YouTube. J Pediatr Adolesc Gynecol 2019 Apr;32(2):153-157. [doi: 10.1016/j.jpag.2018.11.001] [Medline: 30445163]
52. Erdem H, Sisik A. The Reliability of Bariatric Surgery Videos in YouTube Platform. Obes Surg 2018 Mar 5;28(3):712-716. [doi: 10.1007/s11695-017-2911-3] [Medline: 28875396]
53. Faasse K, Chatman CJ, Martin LR. A comparison of language use in pro- and anti-vaccination comments in response to a high profile Facebook post. Vaccine 2016 Nov 11;34(47):5808-5814. [doi: 10.1016/j.vaccine.2016.09.029] [Medline: 27707558]
54. Fullwood MD, Kecojevic A, Basch CH. Examination of YouTube videos related to synthetic cannabinoids. Int J Adolesc Med Health 2016 Aug 17;30(4) [FREE Full text] [doi: 10.1515/ijamh-2016-0073] [Medline: 27639268]
55. Venkatraman A, Garg N, Kumar N. Greater freedom of speech on Web 2.0 correlates with dominance of views linking vaccines to autism. Vaccine 2015 Mar 17;33(12):1422-1425. [doi: 10.1016/j.vaccine.2015.01.078] [Medline: 25665960]
56. Gimenez-Perez G, Robert-Vila N, Tomé-Guerreiro M, Castells I, Mauricio D. Are YouTube videos useful for patient self-education in type 2 diabetes? Health Informatics J 2020 Mar 29;26(1):45-55 [FREE Full text] [doi: 10.1177/1460458218813632] [Medline: 30488752]
57. Goobie GC, Guler SA, Johannson KA, Fisher JH, Ryerson CJ. YouTube Videos as a Source of Misinformation on Idiopathic Pulmonary Fibrosis. Annals ATS 2019 May;16(5):572-579. [doi: 10.1513/annalsats.201809-644oc]
58. Guidry JP, Jin Y, Orr CA, Messner M, Meganck S. Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. Public Relations Review 2017 Sep;43(3):477-486. [doi: 10.1016/j.pubrev.2017.04.009]
59. Guidry J, Jin Y, Haddad L, Zhang Y, Smith J. How Health Risks Are Pinpointed (or Not) on Social Media: The Portrayal of Waterpipe Smoking on Pinterest. Health Commun 2016;31(6):659-667. [doi: 10.1080/10410236.2014.987468] [Medline: 26512916]
60. Guidry JPD, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. Vaccine 2015 Sep 22;33(39):5051-5056. [doi: 10.1016/j.vaccine.2015.08.064] [Medline: 26319742]
61. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013 Sep 06;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]

62. Harris JK, Duncan A, Men V, Shevick N, Krauss MJ, Cavazos-Rehg PA. Messengers and Messages for Tweets That Used #thinspo and #fitspo Hashtags in 2016. Prev Chronic Dis 2018 Jan 04;15:E01 [FREE Full text] [doi: 10.5888/pcd15.170309] [Medline: 29300696]
63. Haymes AT, Harries V. 'How to stop a nosebleed': an assessment of the quality of epistaxis treatment advice on YouTube. J Laryngol Otol 2016 Aug;130(8):749-754. [doi: 10.1017/S0022215116008410] [Medline: 27345303]
64. Helmi M, Spinella MK, Seymour B. Community water fluoridation online: an analysis of the digital media ecosystem. J Public Health Dent 2018 Sep 30;78(4):296-305. [doi: 10.1111/jphd.12268] [Medline: 29603251]
65. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017 Jun 22;35(29):3621-3638 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.052] [Medline: 28554500]
66. Katsuki T, Mackey TK, Cuomo R. Establishing a Link Between Prescription Drug Abuse and Illicit Online Pharmacies: Analysis of Twitter Data. J Med Internet Res 2015 Dec 16;17(12):e280 [FREE Full text] [doi: 10.2196/jmir.5144] [Medline: 26677966]
67. Keelan J, Pavri V, Balakrishnan R, Wilson K. An analysis of the Human Papilloma Virus vaccine debate on MySpace blogs. Vaccine 2010 Feb 10;28(6):1535-1540. [doi: 10.1016/j.vaccine.2009.11.060] [Medline: 20003922]
68. Keim-Malpass J, Mitchell EM, Sun E, Kennedy C. Using Twitter to Understand Public Perceptions Regarding the #HPV Vaccine: Opportunities for Public Health Nurses to Engage in Social Marketing. Public Health Nurs 2017 Jul 06;34(4):316-323. [doi: 10.1111/phn.12318] [Medline: 28261846]
69. Kim R, Park H, Kim H, Kim A, Jang M, Jeon B. Dry facts are not always inviting: a content analysis of Korean videos regarding Parkinson's disease on YouTube. J Clin Neurosci 2017 Dec;46:167-170. [doi: 10.1016/j.jocn.2017.09.001] [Medline: 28988649]
70. Krauss MJ, Grucza RA, Bierut LJ, Cavazos-Rehg PA. “Get drunk. Smoke weed. Have fun.”: A Content Analysis of Tweets About Marijuana and Alcohol. Am J Health Promot 2016 Nov 17;31(3):200-208. [doi: 10.4278/ajhp.150205-qual-708]
71. Krauss MJ, Sowles SJ, Moreno M, Zewdie K, Grucza RA, Bierut LJ, et al. Hookah-Related Twitter Chatter: A Content Analysis. Prev Chronic Dis 2015 Jul 30;12:E121 [FREE Full text] [doi: 10.5888/pcd12.150140] [Medline: 26226068]
72. Kumar N, Pandey A, Venkatraman A, Garg N. Are video sharing web sites a useful source of information on hypertension? J Am Soc Hypertens 2014 Jul;8(7):481-490. [doi: 10.1016/j.jash.2014.05.001] [Medline: 25064770]
73. Laestadius LI, Wahl MM, Cho YI. #Vapelife: An Exploratory Study of Electronic Cigarette Use and Promotion on Instagram. Subst Use Misuse 2016 Oct 14;51(12):1669-1673. [doi: 10.1080/10826084.2016.1188958] [Medline: 27484191]
74. Leong AY, Sanghera R, Jhajj J, Desai N, Jammu BS, Makowsky MJ. Is YouTube Useful as a Source of Health Information for Adults With Type 2 Diabetes? A South Asian Perspective. Can J Diabetes 2018 Aug;42(4):395-403.e4. [doi: 10.1016/j.jcjd.2017.10.056] [Medline: 29282200]
75. Lewis SP, Knoll AKI. Do It Yourself: Examination of Self-Injury First Aid Tips on YouTube. Cyberpsychol Behav Soc Netw 2015 May;18(5):301-304. [doi: 10.1089/cyber.2014.0407] [Medline: 25965864]
76. Loeb S, Sengupta S, Butaney M, Macaluso JN, Czarniecki SW, Robbins R, et al. Dissemination of Misinformative and Biased Information about Prostate Cancer on YouTube. Eur Urol 2019 Apr;75(4):564-567. [doi: 10.1016/j.eururo.2018.10.056] [Medline: 30502104]
77. Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: content drivers and what they are saying. Am J Infect Control 2013 Jun;41(6):568-570. [doi: 10.1016/j.ajic.2012.10.016] [Medline: 23726548]
78. Martinez LS, Hughes S, Walsh-Buhi ER, Tsou M. "Okay, We Get It. You Vape": An Analysis of Geocoded Content, Context, and Sentiment regarding E-Cigarettes on Twitter. J Health Commun 2018 Jul 06;23(6):550-562. [doi: 10.1080/10810730.2018.1493057] [Medline: 29979920]
79. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying Multiple Data Collection Tools to Quantify Human Papillomavirus Vaccine Communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
80. McNeil K, Brna PM, Gordon KE. Epilepsy in the Twitter era: a need to re-tweet the way we think about seizures. Epilepsy Behav 2012 Feb;23(2):127-130. [doi: 10.1016/j.yebeh.2011.10.020] [Medline: 22134096]
81. Menon D, Chelakkot PG, Sunil D, Lakshmaiah A. A cross sectional review of patient information available in the World Wide Web on CyberKnife: fallacies and pitfalls. Support Care Cancer 2017 Dec 5;25(12):3769-3773. [doi: 10.1007/s00520-017-3807-4] [Medline: 28681125]
82. Merianos AL, Gittens OE, Mahabee-Gittens EM. Depiction of Health Effects of Electronic Cigarettes on YouTube. J Subst Use 2016 May 09;21(6):614-619 [FREE Full text] [doi: 10.3109/14659891.2015.1118565] [Medline: 28217030]
83. Meylakhs P, Rykov Y, Koltsova O, Koltsov S. An AIDS-denialist online community on a Russian social networking service: patterns of interactions with newcomers and rhetorical strategies of persuasion. J Med Internet Res 2014 Nov 17;16(11):e261 [FREE Full text] [doi: 10.2196/jmir.3338] [Medline: 25403351]
84. Morin C, Bost I, Mercier A, Dozon J, Atlani-Duault L. Information Circulation in times of Ebola: Twitter and the Sexual Transmission of Ebola by Survivors. PLoS Curr 2018 Aug 28;10 [FREE Full text] [doi: 10.1371/currents.outbreaks.4e35a9446b89c1b46f8308099840d48f] [Medline: 30254789]

85. Mueller SM, Jungo P, Cajacob L, Schwegler S, Itin P, Brandt O. The Absence of Evidence is Evidence of Non-Sense: Cross-Sectional Study on the Quality of Psoriasis-Related Videos on YouTube and Their Reception by Health Seekers. J Med Internet Res 2019 Jan 16;21(1):e11935 [FREE Full text] [doi: 10.2196/11935] [Medline: 30664460]
86. Porat T, Garaizar P, Ferrero M, Jones H, Ashworth M, Vadillo MA. Content and source analysis of popular tweets following a recent case of diphtheria in Spain. Eur J Public Health 2019 Feb 01;29(1):117-122. [doi: 10.1093/eurpub/cky144] [Medline: 30084926]
87. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. JMIR Public Health Surveill 2016;2(1):e1 [FREE Full text] [doi: 10.2196/publichealth.5059] [Medline: 27227144]
88. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018 Jun 14;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]
89. Seltzer E, Horst-Martz E, Lu M, Merchant R. Public sentiment and discourse about Zika virus on Instagram. Public Health 2017 Sep;150:170-175. [doi: 10.1016/j.puhe.2017.07.015] [Medline: 28806618]
90. Seymour B, Getman R, Saraf A, Zhang LH, Kalenderian E. When advocacy obscures accuracy online: digital pandemics of public health misinformation through an antifluoride case study. Am J Public Health 2015 Mar;105(3):517-523. [doi: 10.2105/AJPH.2014.302437] [Medline: 25602893]
91. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
92. Teufel M, Hofer E, Junne F, Sauer H, Zipfel S, Giel KE. A comparative analysis of anorexia nervosa groups on Facebook. Eat Weight Disord 2013 Dec 27;18(4):413-420. [doi: 10.1007/s40519-013-0050-y] [Medline: 23888403]
93. Tiggemann M, Churches O, Mitchell L, Brown Z. Tweeting weight loss: A comparison of #thinspiration and #fitspiration communities on Twitter. Body Image 2018 Jun;25:133-138. [doi: 10.1016/j.bodyim.2018.03.002] [Medline: 29567619]
94. Tuells J, Martínez-Martínez PJ, Duro-Torrijos JL, Caballero P, Fraga-Freijeiro P, Navarro-López V. [Characteristics of the Videos in Spanish Posted on Youtube about Human Papillomavirus Vaccines]. Rev Esp Salud Publica 2015;89(1):107-115 [FREE Full text] [doi: 10.4321/S1135-57272015000100012] [Medline: 25946591]
95. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. Int J Public Health 2016 Mar;61(2):249-256. [doi: 10.1007/s00038-016-0791-2] [Medline: 26841895]
96. Waszak PM, Kasprzycka-Waszak W, Kubanek A. The spread of medical fake news in social media – The pilot quantitative study. Health Policy and Technology 2018 Jun;7(2):115-118. [doi: 10.1016/j.hlpt.2018.03.002]
97. Yang Q, Sangalang A, Rooney M, Maloney E, Emery S, Cappella JN. How Is Marijuana Vaping Portrayed on YouTube? Content, Features, Popularity and Retransmission of Vaping Marijuana YouTube Videos. J Health Commun 2018 Mar 13;23(4):360-369. [doi: 10.1080/10810730.2018.1448488] [Medline: 29533139]
98. Moreno MA, Ton A, Selkie E, Evans Y. Secret Society 123: Understanding the Language of Self-Harm on Instagram. J Adolesc Health 2016 Jan;58(1):78-84 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.09.015] [Medline: 26707231]
99. Sicilia R, Lo Giudice S, Pei Y, Pechenizkiy M, Soda P. Twitter rumour detection in the health domain. Expert Systems with Applications 2018 Nov;110:33-40. [doi: 10.1016/j.eswa.2018.05.019]
100. Shao C, Ciampaglia GL, Varol O, Yang K, Flammini A, Menczer F. The spread of low-credibility content by social bots. Nat Commun 2018 Nov 20;9(1):4787 [FREE Full text] [doi: 10.1038/s41467-018-06930-7] [Medline: 30459415]
101. Lomborg S, Bechmann A. Using APIs for Data Collection on Social Media. The Information Society 2014 Jul 08;30(4):256-265. [doi: 10.1080/01972243.2014.915276]
102. Fast I, Sørensen K, Brand H, Suggs LS. Social media for public health: an exploratory policy analysis. Eur J Public Health 2015 Feb;25(1):162-166. [doi: 10.1093/eurpub/cku080] [Medline: 24942532]

Abbreviations

API: application programming interface
HPV: human papilloma virus

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by A Jamison, A Benis, E Walsh-Buhi, A Budenz, T Porat, JP Allem; comments to author 26.06.20; revised version received 20.08.20; accepted 30.10.20; published 20.01.21.

Please cite as:
Suarez-Lledo V, Alvarez-Galvez J
Prevalence of Health Misinformation on Social Media: Systematic Review
J Med Internet Res 2021;23(1):e17187
URL: http://www.jmir.org/2021/1/e17187/
doi:10.2196/17187
PMID:33470931

Original Paper

Perceived Factors Influencing the Public Intention to Use E-Consultation: Analysis of Web-Based Survey Data

Miaojie Qi[1], MD; Jiyu Cui[1], MPA; Xing Li[1], MPA; Youli Han[1], PhD

[1] School of Public Health, Capital Medical University, Beijing, China

Corresponding Author:
Youli Han, PhD
School of Public Health
Capital Medical University
No10 Xitoutiao Youanmenwai Fengtai District
Beijing, 100069
China
Phone: 86 13520451080
Email: hanyouli@ccmu.edu.cn

Abstract

Background: Unbalanced distribution of medical resources is becoming a major challenge, particularly in the selection of doctors. e-Consultation could provide patients with more choices of doctors and break the constraints of time and space. However, the acceptance of e-consultation is still poor and the mechanism of adoption is unclear.

Objective: The aim of this study was to identify the factors influencing the public intention to use e-consultation and explore the effect path of the factors and behavior intention.

Methods: The hypotheses of our research model were developed based on the technology acceptance model and perceived risk theory. A web-based survey was conducted by an electronic questionnaire collection platform; this survey that consisted of a 29-item questionnaire with 5-point Likert scales was completed by 934 respondents. Structural equation modeling was used to analyze the data. Item evaluation and reliability, validity, path loading, goodness of fit, and multiple group analysis were used to check the moderation effects.

Results: The standardized factor loadings of the items were between 0.551 and 0.873. The composite reliability of 9 constructs ranged from 0.706 to 0.840. The average variance extracted ranged from 0.387 to 0.640. The fitness indices showed that the collected data fitted well with the research model. Perceived usefulness was the strongest positive factor effecting behavior intention (β=.399, P<.001). Perceived ease of use had a positive effect on behavior intention but it was not statistically significant (β=.117, P=.07) and it had a positive effect on perceived usefulness (β=.537, P<.001). Perceived risk could be well explained by financial risk (β=.972, P<.001), privacy risk (β=.774, P<.001), social risk (β=.871, P<.001), time risk (β=.894, P<0.001), and psychological risk (β=.774, P<.001). Perceived risk had negative effects on perceived usefulness (β=–.375, P<.001) and behavior intention (β=–.297, P<.001). Personal innovativeness had a positive influence on perceived ease of use (β=.241, P<.001) and a slight effect on behavior intention (β=.124, P=.001). Age (χ^2_{58}=133.5, P<.001) and usage experience (χ^2_{58}=82.5, P=.02) had a slight moderation effect on the paths.

Conclusions: Perceived usefulness and perceived risk have significant effects on public intention to use e-consultation. Therefore, platform and manufacturer must improve the function of e-consultation, which will promote the public intention to use e-consultation fundamentally. In order to control the perceived risk of public, government should play an important role in enforcing management of e-consultation markets and approving corresponding medical insurance policies. Besides, personal innovativeness had an effect on behavior intention. Moreover, the paths of factors had some heterogeneity among people with different characteristics. Therefore, it is necessary to adjust the strategies to fit more groups better.

(J Med Internet Res 2021;23(1):e21834) doi:10.2196/21834

KEYWORDS

China; e-consultation; perceived risk; surveys and questionnaires; technology acceptance model

Introduction

Background

In China, patients can directly go to tertiary referral hospitals to consult doctors, as primary care physicians do not have sufficient capacity to deal with complex diseases, which results in overloading of tertiary hospitals and increase in the unmet demands of patients [1,2]. With the outbreak of the COVID-19 pandemic, this overwhelming situation is becoming more prominent. e-Consultation provides a new way to solve this challenging situation between doctors and patients. e-Consultation—an innovative way to address the growing medical demand—allows users to overcome the barriers of space and time to have more possibilities of choosing doctors from the whole country, and it is becoming more widespread nowadays. e-Consultation can be classified into 2 types: inter-physician consultation and inter-patient-and-physician consultation [3]. In this study, we focused on the e-consultation between patients and health care providers and we did not involve telecare, telemonitor, and other eHealth. Specifically, users do not need to consult doctors in person and can obtain medical advice asynchronously after uploading personal illness information on the internet through video and text messaging to doctors [4,5].

Compared with face-to-face consultation, e-consultation has some natural advantages and unavoidable limitations. Specially, users only use words, pictures, and videos to communicate with doctors and are unable to receive a medical examination. Therefore, e-consultation is only used on nonurgent minor ailments now [6]. However, according to 2020 World Health Statistics, most patients develop common and chronic diseases, which means that e-consultation could meet great medical demands. In addition to medical advice on diagnosis and treatment regimen, e-consultation could provide patients with timelier and more convenient care [7,8], reduce cost for patients [7,9,10], and improve equitable access for underserved patients and to specialist care [11,12]. For the health system, e-consultation could improve the efficiency of referrals and face-to-face consultation [13-17] and improve the quality of health care [13,18]. Therefore, e-consultation might be a potential solution for major challenges that our health care system faces today [12]. However, many patients are unwilling to use e-consultation even if their illnesses are not serious because they prefer to see doctors in person [19]. A study showed that the average workload of doctors providing web-based health care services was 0.38 patients in China [20] and the situation that people lack awareness of e-consultation exists in a developed country too [21]. For the large part, users are unwilling to believe in the judgments of the doctors in web-based health care services without seeing doctors and without undergoing a medical examination [22]. Our previous survey also found that patients with prior experience of using e-consultation went to the hospital later for the same disease because they wanted to check if the judgement of the doctors providing web-based health care services was accurate. There are great risks perceived by patients if they follow the e-consultation judgement completely. Therefore, many people stated that with the help of their primary care providers, they can use e-consultation better in order to avoid mistakes [23].

As mentioned above, there are many researches focused on the clinical and socioeconomic effects of e-consultation, but these researches have not explored the process or the acceptance of e-consultation or the barriers and the promoters of e-consultation [24]. The content of e-consultation is significantly different from the other functions of eHealth; therefore, we cannot simply apply the usage mechanism of eHealth into e-consultation. This study can fill this gap effectively. The acceptance of e-consultation is a matter of accepting medical information technology, and the technology acceptance model (TAM) explains the acceptance behavior of information communication technology for individuals well [25]. However, our preliminary research and field investigation showed that perceived risk was a significant factor influencing usage behavior, which was reported in many studies as well. However, perceived risk is always taken as a simple dimension in prior researches, which lead to the lack of specific and accurate guiding effects on reality. This research aims to further decompose the perceived risk dimension comprehensively. Therefore, combining TAM and perceived risk, we reconstructed a new model to explore the acceptance mechanism of e-consultation, and we hope this research would help governments and providers make effective and efficient intervention strategies.

Theoretical Background

TAM

The TAM was proposed by Davis based on the previous theories. TAM focuses on an individual's intention to accept information technology. In TAM, perceived usefulness is defined as the extent to which people believe apps would help them perform their job better. Perceived ease of use is defined as the extent to which people believe using apps would be free of effort. Perceived usefulness and perceived ease of use are the 2 main elements that have influence on the intention to use, and sufficient intention leads to actual usage behavior [25]. TAM has been successfully adapted in many eHealth [26], mobile health [27], mobile management systems [28,29], and web-based medical websites [30].

Perceived Risk Theories

A lot of researches show that perceived risk is a key factor that influences people to use medical innovations [28,31-33]. In the medical field, the public always makes medical decisions uncertainly due to information asymmetry, especially when using some emerging medical products and functions. e-Consultation has not really realized the maturity of technology and the stability of the service mode, which aggravates the uncertainty. As shown in the research that even if patients reported satisfaction and acceptance of e-consultation, they did not express strong interests in participating in this interaction because of medical responsibility and accuracy of disease description [34]. Therefore, we take perceived risk as one of the core dimensions of this study and integrate it with TAM.

Originally, perceived risk illustrates the mechanism of people for accepting new brands in the commercial market. It is a sense of uncertainty caused by consumers' inability to predict the outcome of their purchases. The components of perceived risk includes performance risk, physical risk, financial risk, social risk, and psychological risk [35]. With the development of the perceived risk theory, more components are added into the construct, including time risk [36] and privacy risk [37]. Perceived risk theory holds the view that people try to minimize the perceived risk of behavior rather than to maximize the perceived benefit when making consumption decisions [38]. In our study, performance risk is defined as the possibility of e-consultation not performing as it is designed [35,39]. Physical risk is the chances that e-consultation could result in delays in treatment or in misdiagnoses [35]. We can find that delaying treatment or a misdiagnosis means performing out of control; therefore, we just need to keep one factor between performance risk and physical risk. Financial risk and time risk refer to the possibility that users may face loss of money and time when using e-consultation [35,39]. Social risk is the chances that the use of e-consultation would affect the way others think of the users [35,39]. Psychological risk is the chance that e-consultation would not fit in well with users' self-image or self-concept [35,39]. Privacy risk is the potential loss of control over personal information [35,39].

Personal Innovativeness

Personal innovativeness is defined as the degree to which a person is relatively willing to adopt e-consultation in this study [40]. The relationship between technology and the degree of receptiveness to innovation determines how quickly a person adopts information and communications technology [41]. Personal innovativeness can explain the individual differences in their perception of e-consultation advantages and risks. Individuals with higher innovativeness prefer change and tend to gather more information of the technical products. The positive attitude of the innovator toward products would be promoted by the increased interaction with products, which makes them pay more attention to the advantages of technical products and not worry about products working in the designed way [42]. This viewpoint has also been tested in several researches of mobile health adoption [27,43].

Research Model and Hypotheses

The public can choose any registered doctor on the e-consultation platform with a limited cost. The platforms provide users with all kind of hospital departments with different service levels from different regions. Thus, the public have more access for better consultation services. Through e-consultation, users can receive valuable suggestions easily and quickly. After obtaining enough suggestions, they are able to make and follow health decisions better. These functions of e-consultation are attractive to the users. Besides, if it is easy to learn how to use e-consultation, it means that the public will accept e-consultation easier without much effort. Thus, we propose the following hypotheses based on TAM:

Hypothesis 1: Perceived usefulness will influence behavior intention positively.

Hypothesis 2: Perceived ease of use will influence behavior intention positively.

Hypothesis 3: Perceived ease of use will influence perceived usefulness positively.

If e-consultation provides incorrect suggestions, users would be delayed in accepting correct treatment or they may receive wrong treatment. Loss of performance means a loss of health. Unlike the common consumer behavior, performance risk and physical risk are always perceived by the public together. Therefore, physical risk could be absorbed into performance risk. In the TAM, perceived usefulness reflects the functions of e-consultation as well. High levels of risk perceived by people means that they have a suspicion on the usefulness of e-consultation. Therefore, it is unnecessary to integrate performance risk and physical risk into the model again. Besides, since e-consultation needs users to submit symptoms, medical records, and other personal information, the operation of e-consultation would be a new challenge for the user. Thus, we propose the following hypotheses:

Hypothesis 4a: Financial risk is a component of perceived risk of using e-consultation.

Hypothesis 4b: Privacy risk is a component of perceived risk of using e-consultation.

Hypothesis 4c: Social risk is a component of perceived risk of using e-consultation.

Hypothesis 4d: Time risk is a component of perceived risk of using e-consultation.

Hypothesis 4e: Psychological risk is a component of perceived risk of using e-consultation.

Hypothesis 5: Perceived risk will influence perceived usefulness negatively.

Hypothesis 6: Perceived risk will influence perceived ease of use negatively.

Hypothesis 7: Perceived risk will influence behavior intention negatively.

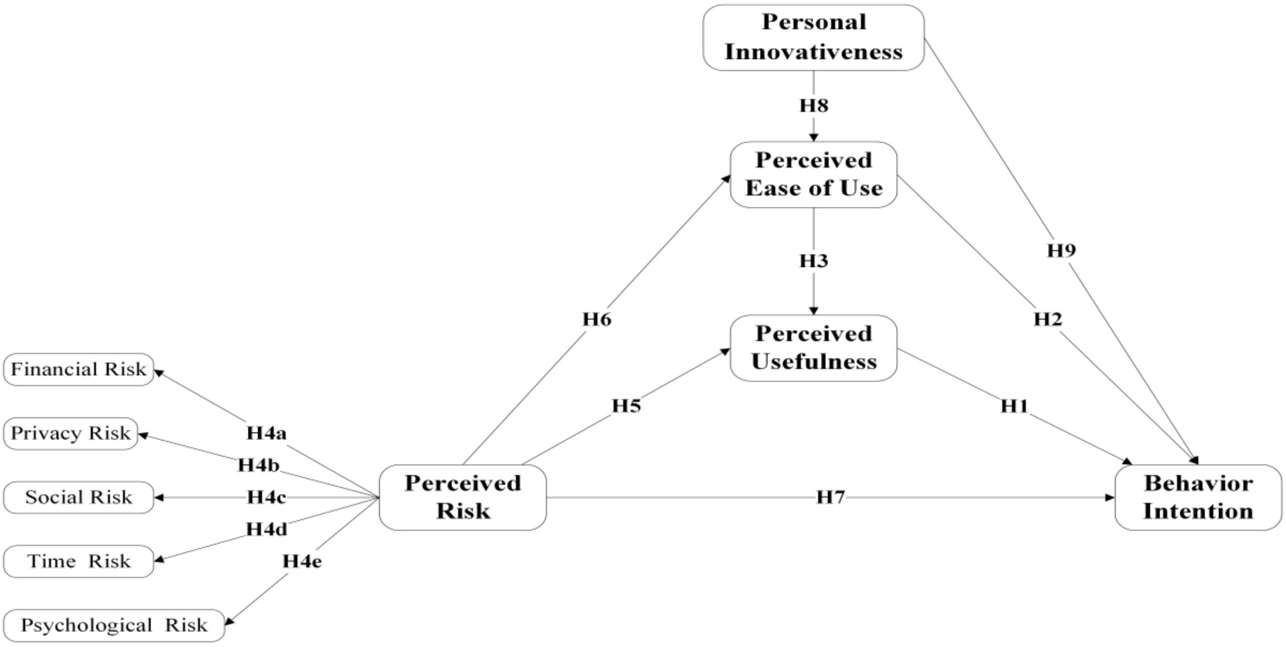
e-Consultation as a combination of information technology and medical services subverts the traditional concept of consultation. Therefore, if people have better innovativeness, they are willing to adopt all kinds of new information technology, including e-consultation. Some other researches show that personal innovativeness also has a direct effect on perceived ease of use. Because people with high level of innovativeness have richer experience in using emerging products, they would think the operation of e-consultation is less difficult. Thus, we propose the following hypotheses:

Hypothesis 8: Personal innovativeness will influence perceived ease of use positively.

Hypothesis 9: Personal innovativeness will influence behavior intention positively.

Overall, the research model is showed in Figure 1.

Figure 1. Research model based on the technology acceptance model and perceived risk theory. Personal innovativeness had effects on behavior intention and perceived ease of use. H: hypothesis.



Aim of This Study

The objective of this study was to investigate people's actual usage of e-consultation and their characteristics. Moreover, based on the TAM integrating with perceived risk and personal innovativeness theory, a questionnaire survey was used to explore the relationships and paths of the factors that influence people's intention to use e-consultation.

Methods

Study Design

All survey items were adopted from previous studies related to eHealth and health information technology. The first version of the questionnaire was directly translated from English to Chinese by a group of researchers. Items were reasonably changed to adapt to the e-consultation. Then, the second version of the questionnaire was completed after 2 rounds of experts' discussions on the first questionnaire. The experts consisted of 2 college professors, 7 staffs from an eHealth company, and 7 doctors with e-consultation using experience. Some items were added or removed or replaced according to the suggestions of experts. The third modification of the questionnaire was completed after a preliminary survey of 222 students majoring in health management from Capital Medical University. Some items were removed or changed to ensure the reliability and validity of the questionnaire. In the end, back translation was performed from Chinese to English by another qualified translator. The final items (Table 1, [44-46]) were measured with a 5-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5). The final questionnaire consisted of 2 parts. The first part was the demographic information of the respondents. The second part, which includes the items for constructs, was designed to measure the respondents' perception on each item.

Table 1. Measurement items of the constructs.

Construct	Item
PU[a] [25,44,45]	
PU1	Using e-consultation would make it easier to consult a specialist or a certified doctor.
PU2	Using e-consultation enables me to understand my disease and treatment recommendation more quickly.
PU3	Using e-consultation facilitates complete communication with doctor.
PU4	Using e-consultation enables me to know more about disease prevention and management.
PU5	Using e-consultation enables me to make better treatment-related decisions.
PU6	I find it easy to obtain information on e-consultation.
PEU[b] [25,44,45]	
PEU1	Learning to use e-consultation is easy for me.
PEU2	In e-consultation, my doctor talks to me clearly and helps me understand my situation appropriately.
PEU3	Using e-consultation would not require much mental effort.
PEU4	It is easy for me to become skillful at using e-consultation.
FR[c] [31,35,39]	
FR1	e-Consultation is not effective and is a waste of money.
FR2	e-Consultation may make me spend extra money in case of a misdiagnosis, leading to delayed correct treatment.
FR3	Using e-consultation may lead to potential fraud.
PRR[d] [35,39]	
PRR1	After using e-consultation, my personal information may be leaked.
PRR2	After using e-consultation, my personal information may be used without my knowledge.
PRR3	After using e-consultation, my illness information may be found by others around me.
SR[e] [31,35,39]	
SR1	If I use e-consultation, it would negatively affect the way others think of me.
SR2	If I use e-consultation, my friends and relatives would think less highly of me.
TR[f] [31,39]	
TR1	e-Consultation may be a waste of time because it is not effective.
TR2	e-Consultation may be a waste of time because of wrong diagnoses or treatments.
PSR[g] [31,35,39]	
PSR1	e-Consultation is not my traditional way to consult doctors, which would lead to psychological issues
PSR2	I am unable to communicate with doctors face-to-face thereby leading to psychological issues
PSR3	I am worried that I cannot describe my disease symptoms correctly when using e-consultation.
PI[h] [40,45,46]	
PI1	I often follow new information technologies with interest.
PI2	If I hear about a new information technology, I would look for ways to experiment with it.
PI3	Among my peers, I am usually the first to try out new information technologies.
BI[i] [25]	
BI1	I intend to use e-consultation.
BI2	I intend to use more e-consultation.
BI3	I predict that I will use e-consultation.

[a] PU: perceived usefulness.

[b] PEU: perceived ease of use.

[c] FR: financial risk.

[d] PRR: privacy risk.

[e] SR: social risk.

[f] TR: time risk.

[g] PSR: psychological risk.

[h] PI: personal innovativeness.

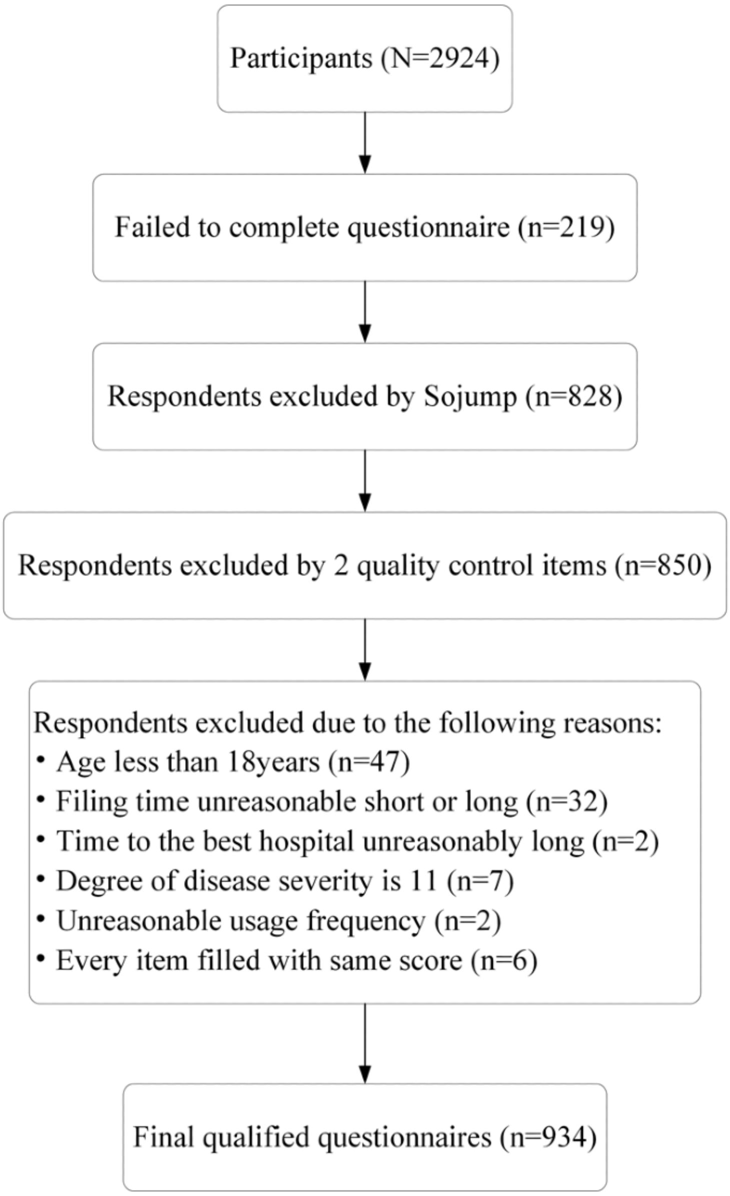
[i] BI: behavior intention.

Data Collection

With the development of information technology, internet protocol restriction, and real-name system, the data quality of web-based surveys meets the requirements of scientific researches. As mentioned above, the users of e-consultation are mainly concentrated in young and middle-aged groups, and the middle-aged group is more willing to accept web-based questionnaires. Therefore, a web-based survey was conducted by Sojump in this research. Sojump is an e-survey company [47], which has 2.6 million samples with all kinds of social demographic characteristics, and unqualified objects can be excluded based on the purpose of the study. The questionnaire was announced on Sojiangwang [48] until the required population was reached. The Sojiangwang is a platform belonging to Sojump, in which all kinds of people can register in. The Sojiangwang asks every registrant to upload the real identity information and audit the identity information. In this platform, the registrant can see all the questionnaires when they meet the included standard of the questionnaires. All different questionnaires would be named with a unified format: "questionnaire + number." In order to ensure the quality of the survey, Sojump uses a series of logical and common sense items to eliminate the halfhearted respondents, and 828 respondents were excluded by this way in our study. Besides, we also designed 2 items to screen the poor-quality questionnaires. The first item listed in the demographic information survey is "e-Consultation could provide surgical and pathological examination services" and the respondents were asked to choose "disagree." The second item listed in the construct is "Now, e-consultation could provide diagnosis and treatment for all kind of diseases" and the respondents were asked to choose "strongly disagree."

At the beginning of the electronic questionnaire, the following information was given first: the purpose of the questionnaire, information and instructions regarding the questionnaire, assurance of proper handling of personal information, and the name of the research institution. The questionnaire link provided on the website (Sojiangwang) could not be copied. After users filled in the questionnaire through the link, the link was removed from the list and could not be used repeatedly. We have reported the results of this survey following the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist, which can be found in Multimedia Appendix 1. The data were collected from March 30, 2020 to April 4, 2020. A total of 2924 participants were involved in this survey, and 934 respondents' questionnaires reached the inclusion criteria of the survey (Figure 2). This study was approved by the ethics committee of the Capital Medical University (number Z2019SY017).

Figure 2. Sampling procedure.



As shown in Table 2, female respondents were more than male respondents. Most respondents were between 18 and 45 years of age (861/934, 92.2%). The education level of the respondents was good because only 21.7% (203/934) of the respondents had a lower level of education than bachelor’s degree. Most respondents were from the more developed eastern region (559/934, 59.9%) and urban region (797/934, 85.4%). The level of income and the access to medical resources were relatively average. Approximately 67.5% (630/934) of the respondents had used e-consultation, of which 80.1% (505/630) had used e-consultation 5 times or less last year. The aims of using e-consultation include helping themselves (352/630, 55.9%) and others (278/630, 44.1%). Both serious and minor diseases could be the subject of e-consultation.

Table 2. Geographical characteristics of the respondents (N=934).

Characteristics	n (%), Value
Gender	
Male	379 (40.6)
Female	555 (59.4)
Age (years)	
18-25	239 (25.6)
26-35	463 (49.6)
36-45	159 (17.0)
>46	73 (7.8)
Education	
Middle school or lower	13 (1.4)
High school	52 (5.6)
Three-year college	138 (14.7)
Bachelor	660 (70.7)
Master or higher	71 (7.6)
Residence	
Rural	137 (14.6)
Urban	797 (85.4)
Location	
Eastern	559 (59.9)
Midregion	164 (17.5)
Western	152 (16.2)
Northern	59 (6.3)
Average annual income (¥, US $1=¥6.475)	
0-10,000	297 (31.8)
11,000-20,000	348 (37.3)
21,000-30,000	174 (18.6)
>30,000	115 (12.3)
Time to the best hospital in district and county (minutes)[a]	
1-10	281 (30.1)
11-20	289 (30.9)
21-30	227 (24.3)
>30	137 (14.7)
Usage experience	
Used	630 (67.5)
Unused	304 (32.5)
Usage frequency last year	
0-2	215 (34.1)
3-5	290 (46.0)
>5	125 (19.9)
Use e-consultation for whom	
Myself	352 (55.9)
Others	278 (44.1)

Characteristics	n (%), Value
Disease severity[b]	
1-5	337 (53.5)
6-10	293 (46.5)

[a] Measures the accessibility of high quality medical resources.

[b] Users themselves assessed the severity of last disease consulted by e-consultation using a 5-point Likert scale ranging from “strongly not serious” (1) to “strongly serious” (10).

Data Analysis

SPSS 20.0 (IBM Corp) was used to analyze the descriptive statistics of respondents’ demographic characteristics and the Cronbach α of the constructs. Amos 24.0 (IBM Corp) was used to evaluate items, measurements, and structural models. Confirmatory factor analysis of the measurement model was used to evaluate the structural model’s path effects, significance, goodness of fit, and moderation effects. Composite reliability and average variance extracted were adopted to evaluate construct reliability and validity.

Results

Measurement Model Testing

The results of reliability and validity are shown in Table 3. The composite reliability, Cronbach α of construct, was greater than the recommended value of .7, and except for financial risk, perceived usefulness, and perceived ease of use, the average variance extracted of constructs was higher than 0.5 [49]. In order to assure the availability of the model, we excluded the low loading items (PU3, PU4) to construct model 2 for testing the result of model 1. We found that although the average variance extracted was promoted a little in model 2, the model fit indices had no substantial improvement (Table 3). Further, the path effects had no substantive difference between model 1 and model 2. Besides, some researches showed that all factors fulfilled a weak or strong validity because factor loadings were statistically significant and the coefficients of path were substantial [50]. Therefore, it was reasonable to accept the results of model 1. As shown in Table 4, the collected data fit well with the research model [51]. The χ^2/df (1111.9/363) of model 1 was 3.1 and was lower than 5. The root mean square error of approximation was 0.047 and was lower than 0.05. The goodness of fit index, comparative fit index, normed fit index, Tucker-Lewis index, and incremental fit index were greater than 0.9 and the adjusted goodness of fit index was 0.9.

Table 3. Item loading and validity.

Construct/Item	Factor loading		Cronbach α		Composite reliability		Average variance extracted	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
PU[a]			.754	.740	0.790	0.740	0.387	0.416
PU1	0.664	0.665						
PU2	0.652	0.651						
PU3	0.577	—[b]						
PU4	0.551	—						
PU5	0.640	0.639						
PU6	0.640	0.625						
PEU[c]			.743	.743	0.747	0.748	0.425	0.426
PEU1	0.623	0.629						
PEU2	0.665	0.658						
PEU3	0.645	0.646						
PEU4	0.673	0.676						
FR[d]			.744	.744	0.745	0.745	0.495	0.494
FR1	0.648	0.647						
FR2	0.734	0.734						
FR3	0.725	0.725						
PRR[e]			.829	.829	0.840	0.840	0.640	0.640
PRR1	0.867	0.867						
PRR2	0.873	0.874						
PRR3	0.638	0.637						
SR[f]			.702	.702	0.706	0.706	0.547	0.547
SR1	0.789	0.789						
SR2	0.687	0.687						
TR[g]			.761	.761	0.761	0.761	0.614	0.614
TR1	0.786	0.786						
TR2	0.781	0.781						
PSR[h]			.754	.754	0.768	0.768	0.529	0.529
PSR1	0.813	0.813						
PSR2	0.766	0.765						
PSR3	0.583	0.583						
PI[i]			.745	.745	0.753	0.753	0.509	0.509
PI1	0.575	0.575						
PI2	0.792	0.792						
PI3	0.754	0.754						
BI[j]			.790	.790	0.789	0.789	0.556	0.556
BI1	0.768	0.768						
BI2	0.768	0.768						
BI3	0.698	0.697						

[a] PU: perceived usefulness.

[b] Not available.

[c] PEU: perceived ease of use.

[d] FR: financial risk.

[e] PRR: privacy risk.

[f] SR: social risk.

[g] TR: time risk.

[h] PSR: psychological risk.

[i] PI: personal innovativeness.

[j] BI: behavior intention.

Table 4. Research model fit.

Fit index	Value (χ^2/*df*)	GFI[a]	AGFI[b]	RMSEA[c]	CFI[d]	NFI[e]	TLI[f]	IFI[g]
Recommended value	<5	>0.9	>0.9	<0.05	>0.9	>0.9	>0.9	>0.9
Value in model 1	3.1 (1111.9/363)	0.917	0.900	0.047	0.932	0.903	0.924	0.933
Value in model 2	3.2 (1002.4/310)	0.919	0.901	0.049	0.934	0.907	0.925	0.934

[a] GFI: goodness of fit index.

[b] AGFI: adjusted goodness of fit index.

[c] RMSEA: root mean square error of approximation.

[d] CFI: comparative fit index.

[e] NFI: normed fit index.

[f] TLI: Tucker-Lewis index.

[g] IFI: incremental fit index.

Structural Model Testing

The judgments of hypotheses based on the SEM results are shown in Table 5. The judgments of model 1 and model 2 exhibited the same results and the standardized factor loadings of path were very closed. Perceived ease of use had no statistically significant effect on behavior intention (β=.117, *P*=.07; β1=.104, *P*=.13). Perceived usefulness had a positive effect on behavior intention (β=.399, *P*<.001; β1=.431, *P*<.001) and was the strongest positive factor of behavior intention. Perceived ease of use had a positive effect on perceived usefulness (β=.537, *P*<.001; β1=.530, *P*<.001). Perceived risk could be well explained by financial risk (β=.972, *P*<.001; β1=.973, *P*<.001), privacy risk (β=.774, *P*<.001; β1=.774, *P*<.001), social risk (β=.871, *P*<.001; β1=.870, *P*<.001), time risk (β=.894, *P*<.001; β1=.894, *P*<.001), and psychological risk (β=.774, *P*<.001; β1=.774, *P*<.001). Among the components, the effect of financial risk was the strongest and that of social risk was the weakest. Perceived risk had negative effects on perceived usefulness (β=–.375, *P*<.001; β1=–.399, *P*<.001) and behavior intention (β=–.297, *P*<.001; β1=–.275, *P*<.001). Personal innovativeness had a positive influence on perceived ease of use (β=.241, *P*<.001; β1=.242, *P*<.001). Compared with other factors, personal innovativeness had a slight effect on behavior intention (β=.124, *P*=.001; β1=.123, *P*=.001).

Table 5. Results of hypothesis testing.

Hypothesis	Path	β[a]	*P* value	Judgement of model 1	β1[b]	*P* value	Judgement of model 2
H1	PU[c]→BI	.399	<.001	Accepted	.431	<.001	Accepted
H2	PEU[d]→BI	.117	.07	Rejected	.104	.13	Rejected
H3	PEU→PU	.537	<.001	Accepted	.530	<.001	Accepted
H4a	PR[e]→FR[f]	.972	<.001	Accepted	.973	<.001	Accepted
H4b	PR→PRR[g]	.774	<.001	Accepted	.774	<.001	Accepted
H4c	PR→SR[h]	.537	<.001	Accepted	.536	<.001	Accepted
H4d	PR→TR[i]	.894	<.001	Accepted	.894	<.001	Accepted
H4e	PR→PSR[j]	.871	<.001	Accepted	.870	<.001	Accepted
H5	PR→PU	–.375	<.001	Accepted	–.399	<.001	Accepted
H6	PR→PEU	–.491	<.001	Accepted	–.488	<.001	Accepted
H7	PR→BI	–.297	<.001	Accepted	–.275	<.001	Accepted
H8	PI[k]→PEU	.241	<.001	Accepted	.242	<.001	Accepted
H9	PI→BI[l]	.124	.001	Accepted	.123	.001	Accepted

[a] β: standardized factor loading of model 1.

[b] β1: standardized factor loading of model 2.

[c] PU: perceived usefulness.

[d] PEU: perceived ease of use.

[e] PR: perceived risk.

[f] FR: financial risk.

[g] PRR: privacy risk.

[h] SR: social risk.

[i] TR: time risk.

[j] PSR: psychological risk.

[k] PI: personal innovativeness.

[l] BI: behavior intention.

Moderation Effect Testing

We further tested the moderating effects of geographical characteristics by multiple-group analysis [52,53]. In order to simplify the data analysis, the total sample was reclassified into 2 subgroups (Table 6). First, to screen the factors with moderation effects from characteristics, we constrained the measurement weights, structural weights, structural covariances, structural residua, and measurement residua of the subgroup model to construct parameter constraints models. If the results of the constraints model and the unconstrained model were significantly different, it indicated that the paths between subgroups are the factors that might have a moderation effect. As showed in Table 6, age (χ^2_{58}=133.5, *P*<.001), income (χ^2_{58}=85.6, *P*=.01), and usage experience (χ^2_{58}=82.5, *P*=.02) might have moderation effects.

Table 6. Dichotomous geographical characteristics of the respondents (N=934).

Characteristics	n (%), Value	χ^2 *(df)*	*P* value
Gender		43.3 (58)	.93
Male	379 (40.6)		
Female	555 (59.4)		
Age (years)		133.5 (58)	<.001
18-30	464 (49.7)		
>30	470 (50.3)		
Education		25.7 (58)	>.99
Three-year college or lower	203 (21.7)		
Bachelor or high	731 (78.3)		
Residence		47.8 (58)	.83
Rural	137 (14.6)		
Urban	797 (85.4)		
Location		38.2 (58)	.98
Eastern	559 (59.9)		
Not eastern	375 (40.1)		
Income (¥, US $ 1 =¥6.475)		85.6 (58)	.01
0-10,000	297 (31.8)		
>10,000	637 (68.2)		
Time to the best hospital (minutes)		35.7 (58)	.99
1-20	570 (61.0)		
>21	364 (39.0)		
Usage experience		82.5 (58)	.02
Used	630 (67.5)		
Unused	304 (32.5)		
Usage frequency last year		49.4 (58)	.78
0-2	215 (34.1)		
>2	415 (65.9)		
Use e-consultation for whom		44.1 (58)	.91
Myself	352 (55.9)		
Others	278 (44.1)		
Disease severity		70.5 (58)	.13
1-5	337 (53.5)		
6-10	293 (46.5)		

Second, we estimated the path loadings and the critical ratios for differences of each subgroup (Table 7). If the absolute value of the critical ratio was lower than 1.96, there would be a significant difference between the paths of the 2 subgroups. Compared with the older subgroup (β_{age2}=.235, *P*=.02), it is estimated that perceived usefulness has more positive effect on behavior intention in the younger subgroup (β_{age1}=.537, *P*<.001). The path loading of hypothesis 9 was not significant in the older subgroup (β_{age1}=.054, *P*=.34). Income had no significant moderation influence on the research model. Besides, the usage experience only had some influence on path coefficient. It is shown that the path loading of perceived ease of use to behavior intention has a significant difference, but the coefficients were very close ($\beta_{experience1}$=.532; $\beta_{experience2}$=.534). The path loading of personal innovativeness to perceived ease of use in the used group ($\beta_{experience1}$=.149, *P*=.008) was lower than that of the unused group ($\beta_{experience1}$=.327, *P*<.001). We found that there was no substantial difference in the usage mechanism whether or not the public used e-consultation.

Table 7. Multiple group analysis.

Hypothesis (H)	β_{age1}[a]	β_{age2}[b]	CR[c]	$\beta_{income1}$[d]	$\beta_{income2}$[e]	CR	$\beta_{experience1}$[f]	$\beta_{experience2}$[g]	CR
H1	.537***	.235 (.022)	–2.221**	.346***	.432***	0.203	.498***	.345***	0.02
H2	.09 (.328)	.172 (.057)	0.556	.115 (.251)	.111 (.193)	–0.030	.045 (.630)	.137 (.162)	0.629
H3	.464***	.587***	1.025	.554***	.538***	0.571	.532***	.534***	–2.043**
H4a	.957***	.987***	—[h]	.942***	.980***	—	.980***	.944***	—
H4b	.741***	.795***	–0.226	.754***	.766***	–1.338	.769***	.688***	0.364
H4c	.442***	.630***	0.511	.515***	.512***	–1.729	.557***	.376***	0.625
H4d	.847***	.943***	–0.694	.800***	.932***	–1.507	.874***	.851***	1.087
H4e	.842***	.899***	–1.274	.821***	.880***	–1.632	.880***	.745***	0.787
H5	–.391***	–.367***	–0.294	–.297***	–.390***	–0.252	–.393***	–.314***	0.306
H6	–.534***	–.456***	0.022	–.408***	–.501***	0.538	–.490***	–.318***	–0.513
H7	–.221 (.004)	–.371***	–1.704	–.323***	–.272***	1.386	–.247***	–.255 (.001)	–1.193
H8	.203 (.002)	.284***	1.831	.258 (.002)	.229***	–0.704	.149 (.008)	.327***	2.785**
H9	.054 (.344)	.186***	2.159**	.197 (.004)	.097 (.039)	–1.47	.125 (.016)	.127 (.063)	0.594

[a] β_{age1}: standardized factor loading of age from 18 years to 30 years.

[b] β_{age2}: standardized factor loading of age over 30 years.

[c] CR: critical ratios for differences.

[d] $\beta_{income1}$: standardized factor loading of income below ¥100,000 per year; US $1=¥6.475.

[e] $\beta_{income2}$: standardized factor loading of income over ¥100,000 per year.

[f] $\beta_{experience1}$: standardized factor loading of the used.

[g] $\beta_{experience1}$: standardized factor loading of the unused.

[h] Not available due to fixed parameter.

**P<.05.

***P<.001.

Discussion

Principal Results

Our study found that perceived usefulness is one of the most important determinants of individuals' intention to use e-consultation, which is similar to that reported in most related studies on the acceptance of information communication technology [27,29,30]. Even in subgroups with different characteristics, the direction and significance of path loading were not changed. Our results indicate that promoting the function of e-consultation is a key to attract the public to use it because the higher perceived usefulness means the public have more trust in the ability and integrity of doctors and platforms [24]. Compared with face-to-face consultation, e-consultation could only be used to diagnose common and chronic diseases lacking necessary medical examinations and the supporting treatment system, but the text suggestions from the specialists are still important for the public. In particular, in some special cases (eg, COVID-19 pandemic), e-consultation could not only achieve the goal of public isolation but also meet the patients' demand of medical services. Although e-consultation cannot be a complete substitute for face-to-face consultations, it may serve as an entry level consultation after integrated into the face-to-face consultation [54]. In addition to the service ability of e-consultation, technical difficulties, including substandard signal construction, virtual device, and video equipment, would significantly weaken people's evaluation on the usefulness of e-consultation [55]. Therefore, while improving the functionality, providers should also pay attention to improve the facilitation condition of e-consultation.

Different from the related researches, our research shows that perceived ease of use has no effect on the behavior intention but it has a strong effect on perceived usefulness. Normally, since the information communication technology products are used in the professional field, learning to use these products is a challenge for users. However, e-consultation just needs users to interact with doctors on the internet by using a personal computer or a smartphone. With the popularity of smartphones in China and worldwide [56], it is reasonable to believe that the public can easily learn how to operate e-consultation; therefore, the ease of use no longer plays a role in the promotion. As shown in a survey of 947 respondents, less than 20% of the people think that the reason they do not use e-consultation is that they are not skillful enough to complete the operation [57]. Another possible explanation is related to the characteristics of the respondents. Perceived ease of use comprises ease of operation, understanding, and expression. In our survey, most of the respondents had high educational backgrounds and were young, which leads to a stronger understanding ability, thereby leading to no significant relationship between perceived ease of use and behavior intention.

Personal innovativeness has a direct effect on perceived ease of use and behavior intention, which is consistent with the findings of a previous study [45]. Although more and more high-tech products are emerging, the public are always keen to try the popular products rather than the new ones. Therefore, it is essential to strengthen the publicity of e-consultation for the public. Of course, formatting innovativeness is a complex and long process [58]; therefore, finding the innovative individual might be a better choice. In the promotion of e-consultation, e-consultation providers should offer advanced services for the innovative and stable services for the common.

Perceived risk has a significantly negative effect on behavior intention. Because of health issues, the public would take risks more seriously and perform risk aversion [59]. Uncertainty and information asymmetry are typical features of medical services, which always leads patients with common diseases to fail in selecting the most effective services (primary health care). They would prefer to go to a tertiary hospital for the minimization of medical risk instead of the maximization of utility [60]. Different from other results that risk influence intention [29,38] or attitude [37] directly, our results show that perceived risk weakens not only behavior intention but also perceived usefulness and perceived ease of use directly. It shows that if we do not control the risk of e-consultation, even if e-consultation could provide services for more diseases, the public would reduce the evaluation and the intention of e-consultation. In addition, perceived usefulness contains performance risk and physical risk and perceived ease of use reflects the risk of operation. It indicates that performance risk, physical risk, and operation risk are components of perceived risk, which is consistent with the theoretical hypothesis and our previous surveys. However, perceived ease of use has no effect on behavior intention. Therefore, we need to take note that ease of operation may not promote the usage intention, but the difficulty of operation may reduce the intention.

There are many studies on the barriers in e-consultation, but these only explored the objective and external factors, for example, signal coverage, equipment, and characteristics of the patients. Even if some qualitative studies ask users' subjective evaluation, the final results are not comprehensive [61]. Our study made up for some gaps in these researches and found that among the other components, financial risk and time risk were the most considered by people. In China, the e-consultation platforms provide free and paid consultation services for users. The choices of doctors and number of questions would be limited in the free services and the cost of paid services cannot be submitted to medical insurance. Therefore, if e-consultation is not effective, the cost and time of using e-consultation will be wasted altogether. Compared with the indirect costs [62] saved by e-consultation, the opportunity cost of e-consultation is more valued by the public. Besides, not all patients believed that e-consultations could play the role in reducing the time to access specialists' advice [63]. Therefore, it is important to strengthen the connection between e-consultation and offline treatment and include the cost of e-consultation into the medical insurance system. These are the 2 keys to promote e-consultation use.

Many studies have found that the key barriers of using e-consultation for patients are privacy concerns and security of their data [5,19,64]. The path of perceived risk and privacy risk showed that providers should strengthen the construction of e-consultation information systems. Privacy disclosure has been a big problem in the medical field [65,66]. Users worry about not only the illegal disclosure but also the exposure to their family members with some special diseases such as mental illness [67]. We should appeal to the government to make related laws and strengthen the supervision of the operation of the e-consultation platform, and then, the public would upload their personal medical information during e-consultation.

e-Consultation, a new consultation model, has been in China for less than 20 years, and the public have little detailed knowledge about it; therefore, the public cannot get used to this kind of non–face-to-face consultation quickly, which has aroused the public's attention to psychological risk. Besides, body language is often accompanied by patients' expression, but the text-based e-consultation cannot reveal the body language, which can easily cause anxiety about the incomplete expression for patients [68]. Real-time video calls could alleviate this problem to some extent, but it is not applicable to all patients because of limitations in different video equipment. Of course, since the public with general health literacy are often unable to describe the uncommon disease symptoms and feelings correctly [69], they would worry about their medical behavior in e-consultation. Therefore, it would be necessary to improve the public's health literacy to decrease the psychological risk. In fact, the lack of people's health literacy is a long-standing problem and it is difficult to be overcome completely [70]. We need to cooperate with certain auxiliary ways to assist the public to use e-consultation, among which keeping a special receptionist [71] may be a good solution.

Although social risk is only a minor component of perceived risk, we need to improve the awareness of e-consultation among the public to help them understand it correctly. As mentioned above, the public know less about e-consultation; therefore, there is no effective consensus on e-consultation in the society. Some researches show that the most prominent reason for nonuse of e-consultation is that the public are not aware of the existence of the service [56,71]. With the improvement of awareness, the public would think it is a reasonable choice to use e-consultation and would not make negative assessments on it.

Our results show that personal innovativeness has an effect on behavior intention for the older population but has no effect for the younger population. A study on users not using e-consultation also showed that age had a moderation effect on behavior intention [57]. We think that for the young, especially between the ages from 18 years to 30 years in our study, their innovativeness generally has a high level; therefore, the path loading of personal innovativeness and behavior intention are not significant. Besides, perceived usefulness has less effect on the behavior intention for the older, because with the increase in age, patients place more emphasis on service attitude and medical process and not just utility [72]. Compared with the age factor, the usage experience has only a slight moderation effect on the usage mechanism. The effect direction of the paths is not changed, which indicates that the mechanisms of factors are consistent between the used and unused. Of course, we can also find that personal innovativeness has more effect on the perceived ease of use for the unused. Therefore, raising the innovativeness of the unused might achieve better effect during the promotion of e-consultation.

Limitations

Our data was collected by a web-based survey; therefore, some selection bias was unavoidable. First, this study showed that 67.5% (630/934) of the respondents were experienced in using e-consultation; however, the usage rate of the students was only 25.7% in our previous survey. These data show that our respondents use mobile devices or computers more frequently. Thus, a higher awareness of e-consultation was observed among these respondents. Second, most respondents came from urban areas (797/934, 85.4%). They might be less willing to use e-consultation because it is easier for them to receive high quality medical resources in the cities. Third, the average age of the respondents was 31 years in our study and 78.3% (731/934) of the respondents had higher education degrees than bachelor’s degree; therefore, our sample may have less medical demand and usage of e-consultation [56]. Besides, the young sample would influence the moderation effect of age. Although it is a better way to survey more people with usage experience through web-based surveys, it is not suitable for all people such as the older adults or the undereducated. Therefore, further offline population-based surveys are necessary, which could be a household survey of residents for small samples with cluster-stratified sampling. In addition, our survey meets the requirements of health care consultations during the COVID-19 pandemic, wherein the public had to stay at home, which might make people have a high intention to use e-consultation.

Conclusions

Our research focuses on the positive and negative factors that influence the public acceptance of e-consultation and supports the use of TAM and perceived risk in explaining public intention to use e-consultation. We found that perceived usefulness and perceived risk are the most important determinants effecting people’s intention to use e-consultation. Therefore, platforms and manufacturers must improve the function of e-consultation, which will promote the public intention to use it fundamentally. Further, to control the perceived risk of public, government should play an important role in enforcing management of e-consultation markets and approving corresponding medical insurance policies. Besides, we found that personal innovativeness has an effect on behavior intention and the path of factors has differences among people with different characteristics to some degree. Therefore, it is necessary to adjust the strategies to adapt to different groups.

Acknowledgments

We thank all experts for offering comments on the design of the questionnaire and Qingdu Meng for editing this manuscript for language. We also thank the Sojump company for assisting in the collection of electronic questionnaires. This work was supported by the Social Science Foundation of Beijing of China (17JDGLA032).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys.
[DOCX File , 22 KB - jmir_v23i1e21834_app1.docx]

References

1. Deng F, Lv J, Wang H, Gao J, Zhou Z. Expanding public health in China: an empirical analysis of healthcare inputs and outputs. Public Health 2017 Jan;142:73-84. [doi: 10.1016/j.puhe.2016.10.007] [Medline: 28057203]
2. Malmivaara A. On decreasing inequality in health care in a cost-effective way. BMC Health Serv Res 2014 Feb 20;14:79 [FREE Full text] [doi: 10.1186/1472-6963-14-79] [Medline: 24552288]
3. Zallman L, Fisher F, Ladner S, Mengistu K, Rapaport B, Bor D, et al. Inter-clinician eConsults without programmatic incentives or requirements: a qualitative study of primary care provider perspectives. Fam Pract 2020 Sep 05;37(4):525-529. [doi: 10.1093/fampra/cmaa016] [Medline: 32112080]
4. Pecina JL, North F. Early e-consultation face-to-face conversions. J Telemed Telecare 2015 Sep 22;22(5):269-276. [doi: 10.1177/1357633x15602634]
5. Mold F, Hendy J, Lai Y, de Lusignan S. Electronic Consultation in Primary Care Between Providers and Patients: Systematic Review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
6. Nijland N, van Gemert-Pijnen JEWC, Kelders S, Brandenburg B, Seydel E. Evaluation of the use of an "ask-the-expert" e-consultation service for support on health-related requests. Stud Health Technol Inform 2010;160(Pt 2):821-825. [doi: 10.1109/etelemed.2010.33] [Medline: 20841800]
7. Zandbelt LC, de Kanter FE, Ubbink DT. E-consulting in a medical specialist setting: Medicine of the future? Patient Educ Couns 2016 May;99(5):689-705. [doi: 10.1016/j.pec.2015.11.005] [Medline: 26597384]

8. Liddy C, Rowan MS, Afkham A, Maranger J, Keely E. Building access to specialist care through e-consultation. Open Med 2013;7(1):e1-e8 [FREE Full text] [Medline: 23687533]
9. Castaneda PR, Duffy B, Andraska EA, Stevens J, Reschke K, Osborne N, et al. Outcomes and safety of electronic consult use in vascular surgery. J Vasc Surg 2020 May;71(5):1726-1732. [doi: 10.1016/j.jvs.2019.08.231] [Medline: 31611112]
10. Liddy C, Keely E. Using the Quadruple Aim Framework to Measure Impact of Heath Technology Implementation: A Case Study of eConsult. J Am Board Fam Med 2018;31(3):445-455 [FREE Full text] [doi: 10.3122/jabfm.2018.03.170397] [Medline: 29743227]
11. Liddy C, Joschko J, Guglani S, Afkham A, Keely E. Improving Equity of Access Through Electronic Consultation: A Case Study of an eConsult Service. Front Public Health 2019;7:279 [FREE Full text] [doi: 10.3389/fpubh.2019.00279] [Medline: 31637228]
12. Anderson D, Porto A, Koppel J, Macri G, Wright M. Impact of Endocrinology eConsults on Access to Endocrinology Care for Medicaid Patients. Telemed J E Health 2020 Nov;26(11):1383-1390. [doi: 10.1089/tmj.2019.0238] [Medline: 32023182]
13. Ong S, Kaushal A, Pariser P, Chan C. An Integrated Kidney Care eConsult Practice Model: Results from the iKinect Project. Am J Nephrol 2019;50(4):262-271. [doi: 10.1159/000502602] [Medline: 31473734]
14. Liddy C, de Man G, Moroz I, Afkham A, Mercer J, Keely E. Effective Integration of an eConsult Service into an Existing Referral Workflow Within a Primary Care Clinic. Telemed J E Health 2020 May;26(5):659-664. [doi: 10.1089/tmj.2019.0014] [Medline: 31414979]
15. Afable MK, Gupte G, Simon SR, Shanahan J, Vimalananda V, Kim EJ, et al. Innovative Use Of Electronic Consultations In Preoperative Anesthesiology Evaluation At VA Medical Centers In New England. Health Aff (Millwood) 2018 Feb;37(2):275-282. [doi: 10.1377/hlthaff.2017.1087] [Medline: 29401018]
16. Wang RF, Trinidad J, Lawrence J, Pootrakul L, Forrest LA, Goist K, et al. Improved patient access and outcomes with the integration of an eConsult program (teledermatology) within a large academic medical center. J Am Acad Dermatol 2020 Dec;83(6):1633-1638. [doi: 10.1016/j.jaad.2019.10.053] [Medline: 31678336]
17. Gilani S, Bommakanti K, Friedman L. Electronic Consults in Otolaryngology: A Pilot Study to Evaluate the Use, Content, and Outcomes in an Academic Health System. Ann Otol Rhinol Laryngol 2020 Feb;129(2):170-174. [doi: 10.1177/0003489419882726] [Medline: 31625409]
18. Tande A, Berbari E, Ramar P, Ponamgi S, Sharma U, Philpot L, et al. Association of a Remotely Offered Infectious Diseases eConsult Service With Improved Clinical Outcomes. Open Forum Infect Dis 2020 Jan;7(1):ofaa003 [FREE Full text] [doi: 10.1093/ofid/ofaa003] [Medline: 31988969]
19. Osman MA, Schick-Makaroff K, Thompson S, Bialy L, Featherstone R, Kurzawa J, et al. Barriers and facilitators for implementation of electronic consultations (eConsult) to enhance access to specialist care: a scoping review. BMJ Glob Health 2019;4(5):e001629 [FREE Full text] [doi: 10.1136/bmjgh-2019-001629] [Medline: 31565409]
20. Li Y, Yan X, Song X. Provision of Paid Web-Based Medical Consultation in China: Cross-Sectional Analysis of Data From a Medical Consultation Website. J Med Internet Res 2019 Jun 03;21(6):e12126 [FREE Full text] [doi: 10.2196/12126] [Medline: 31162129]
21. Ackerman SL, Gleason N, Shipman SA. Comparing Patients' Experiences with Electronic and Traditional Consultation: Results from a Multisite Survey. J Gen Intern Med 2020 Apr;35(4):1135-1142. [doi: 10.1007/s11606-020-05703-7] [Medline: 32076987]
22. de Graaf M, Totte J, Breugem C, van Os-Medendorp H, Pasmans S. Evaluation of the Compliance, Acceptance, and Usability of a Web-Based eHealth Intervention for Parents of Children With Infantile Hemangiomas: Usability Study. JMIR Res Protoc 2013 Dec 17;2(2):e54 [FREE Full text] [doi: 10.2196/resprot.2897] [Medline: 24345450]
23. Joschko J, Liddy C, Moroz I, Reiche M, Crowe L, Afkham A, et al. Just a click away: exploring patients' perspectives on receiving care through the Champlain BASETM eConsult service. Fam Pract 2018 Jan 16;35(1):93-98. [doi: 10.1093/fampra/cmx073] [Medline: 28968806]
24. Li J, Liu M, Liu X, Ma L. Why and When do Patients Use e-Consultation Services? The Trust and Resource Supplementary Perspectives. Telemed J E Health 2018 Jan;24(1):77-85. [doi: 10.1089/tmj.2016.0268] [Medline: 28686084]
25. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008] [Medline: 10.2307/249008]
26. Tsai C. Integrating social capital theory, social cognitive theory, and the technology acceptance model to explore a behavioral model of telehealth systems. Int J Environ Res Public Health 2014 May 07;11(5):4905-4925 [FREE Full text] [doi: 10.3390/ijerph110504905] [Medline: 24810577]
27. Hoque MR. An empirical study of mHealth adoption in a developing country: the moderating effect of gender concern. BMC Med Inform Decis Mak 2016 May 03;16:51 [FREE Full text] [doi: 10.1186/s12911-016-0289-0] [Medline: 27142844]
28. Zhu Z, Liu Y, Che X, Chen X. Moderating factors influencing adoption of a mobile chronic disease management system in China. Inform Health Soc Care 2018 Jan;43(1):22-41. [doi: 10.1080/17538157.2016.1255631] [Medline: 28068149]
29. Jeon E, Park H. Factors affecting acceptance of smartphone application for management of obesity. Healthc Inform Res 2015 Apr;21(2):74-82 [FREE Full text] [doi: 10.4258/hir.2015.21.2.74] [Medline: 25995959]
30. Tang Y, Yang Y, Shao Y. Acceptance of Online Medical Websites: An Empirical Study in China. Int J Environ Res Public Health 2019 Mar 15;16(6):943 [FREE Full text] [doi: 10.3390/ijerph16060943] [Medline: 30884754]

31. Hosoda T, Fraser JR, Kim M, Cheon HJ. Consumer Behavioral Systems to Approach or Avoid Generic Medicine (GM) Consumption in Japan. JAFEB 2018 May 31;5(2):105-118. [doi: 10.13106/jafeb.2018.vol5.no2.105] [Medline: 10.13106/jafeb.2018.vol5.no2.105]
32. Andrews L, Gajanayake R, Sahama T. The Australian general public's perceptions of having a personally controlled electronic health record (PCEHR). Int J Med Inform 2014 Dec;83(12):889-900. [doi: 10.1016/j.ijmedinf.2014.08.002] [Medline: 25200198]
33. Deng Z, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. Int J Med Inform 2017 Sep;105:98-109. [doi: 10.1016/j.ijmedinf.2017.05.014] [Medline: 28750916]
34. Olayiwola JN, Knox M, Dubé K, Lu EC, Woldeyesus T, James IE, et al. Understanding the Potential for Patient Engagement in Electronic Consultation and Referral Systems: Lessons From One Safety Net System. Health Serv Res 2018 Aug;53(4):2483-2502 [FREE Full text] [doi: 10.1111/1475-6773.12776] [Medline: 28940495]
35. Jacoby J, Kaplan L. The Components of Perceived Risk. Association for Consumer Research 1972 Jan 01:1-19. [doi: 10.4337/9781843768708.00081] [Medline: 10.4337/9781843768708.00081]
36. Stone RN, Grønhaug K. Perceived Risk: Further Considerations for the Marketing Discipline. European Journal of Marketing 1993 Apr;27(3):39-50. [doi: 10.1108/03090569310026637]
37. Jarvenpaa SL, Todd PA. Consumer Reactions to Electronic Shopping on the World Wide Web. International Journal of Electronic Commerce 2015 Dec 11;1(2):59-88. [doi: 10.1080/10864415.1996.11518283]
38. Mitchell V. Consumer perceived risk: conceptualisations and models. European Journal of Marketing 1999 Feb;33(1/2):163-195. [doi: 10.1108/03090569910249229]
39. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. International Journal of Human-Computer Studies 2003 Oct;59(4):451-474. [doi: 10.1016/S1071-5819(03)00111-3]
40. Goldsmith RE, Hofacker CF. Measuring consumer innovativeness. JAMS 1991 Jun;19(3):209-221. [doi: 10.1007/bf02726497]
41. Agarwal R, Prasad J. A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Information Systems Research 1998 Jun;9(2):204-215. [doi: 10.1287/isre.9.2.204]
42. Schweitzer F, Van den Hende EA. To Be or Not to Be in Thrall to the March of Smart Products. Psychol Mark 2016 Oct;33(10):830-842 [FREE Full text] [doi: 10.1002/mar.20920] [Medline: 27980356]
43. Chevalking SL, Ben Allouch S, Brusse-Keizer M, Postel MG, Pieterse ME. Identification of Users for a Smoking Cessation Mobile App: Quantitative Study. J Med Internet Res 2018 Apr 09;20(4):e118 [FREE Full text] [doi: 10.2196/jmir.7606] [Medline: 29631988]
44. Lakshmi KB, Rajaram M. Impact of information technology reliance and innovativeness on rural healthcare services: study of dindigul district in Tamilnadu, India. Telemed J E Health 2012 Jun;18(5):360-370. [doi: 10.1089/tmj.2011.0191] [Medline: 22568437]
45. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012 Nov 05;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]
46. Lewis, Agarwal, Sambamurthy. Sources of Influence on Beliefs about Information Technology Use: An Empirical Study of Knowledge Workers. MIS Quarterly 2003;27(4):657. [doi: 10.2307/30036552]
47. Web-based questionnaires. URL: www.wjx.cn [accessed 2019-05-01]
48. Sojiangwang. URL: www.sojiang.com [accessed 2019-05-01]
49. Ye T, Xue J, He M, Gu J, Lin H, Xu B, et al. Psychosocial Factors Affecting Artificial Intelligence Adoption in Health Care in China: Cross-Sectional Study. J Med Internet Res 2019 Oct 17;21(10):e14316 [FREE Full text] [doi: 10.2196/14316] [Medline: 31625950]
50. Koivumäki T, Pekkarinen S, Lappi M, Väisänen J, Juntunen J, Pikkarainen M. Consumer Adoption of Future MyData-Based Preventive eHealth Services: An Acceptance Model and Survey Study. J Med Internet Res 2017 Dec 22;19(12):e429 [FREE Full text] [doi: 10.2196/jmir.7821] [Medline: 29273574]
51. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118] [Medline: 10.1080/10705519909540118]
52. Businelle MS, Kendzor DE, Reitzel LR, Costello TJ, Cofta-Woerpel L, Li Y, et al. Mechanisms linking socioeconomic status to smoking cessation: a structural equation modeling approach. Health Psychol 2010 May;29(3):262-273 [FREE Full text] [doi: 10.1037/a0019285] [Medline: 20496980]
53. Augustus-Horvath CL, Tylka TL. The acceptance model of intuitive eating: a comparison of women in emerging adulthood, early adulthood, and middle adulthood. J Couns Psychol 2011 Jan;58(1):110-125. [doi: 10.1037/a0022129] [Medline: 21244144]
54. Pecina JL, North F. Early e-consultation face-to-face conversions. J Telemed Telecare 2015 Sep 22;22(5):269-276. [doi: 10.1177/1357633x15602634]

55. Abdolahi A, Bull MT, Darwin KC, Venkataraman V, Grana MJ, Dorsey ER, et al. A feasibility study of conducting the Montreal Cognitive Assessment remotely in individuals with movement disorders. Health Informatics J 2016 Jun;22(2):304-311 [FREE Full text] [doi: 10.1177/1460458214556373] [Medline: 25391849]
56. Billieux J. Problematic Use of the Mobile Phone: A Literature Review and a Pathways Model. CPSR 2012 Oct 01;8(4):299-307. [doi: 10.2174/157340012803520522] [Medline: 10.2174/157340012803520522]
57. Nijland N, van Gemert-Pijnen JEWC, Boer H, Steehouder MF, Seydel ER. Increasing the use of e-consultation in primary care: results of an online survey among non-users of e-consultation. Int J Med Inform 2009 Oct;78(10):688-703. [doi: 10.1016/j.ijmedinf.2009.06.002] [Medline: 19625210]
58. Griffin AS. Innovativeness as an emergent property: a new alignment of comparative and experimental research on animal innovation. Philos Trans R Soc Lond B Biol Sci 2016 Mar 19;371(1690):1-9 [FREE Full text] [doi: 10.1098/rstb.2015.0544] [Medline: 26926287]
59. Arrow K. Uncertainty and the welfare economics of medical care. American Economic Association 1963;53(5):941-973. [doi: 10.1016/b978-0-12-214850-7.50028-0] [Medline: 10.1016/b978-0-12-214850-7.50028-0]
60. Peter JP, Tarpey, Sr. LX. A Comparative Analysis of Three Consumer Decision Strategies. J CONSUM RES 1975 Jun;2(1):29. [doi: 10.1086/208613] [Medline: 10.1086/208613]
61. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and Facilitators That Influence Telemedicine-Based, Real-Time, Online Consultation at Patients' Homes: Systematic Literature Review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
62. Kirsh S, Carey E, Aron DC, Cardenas O, Graham G, Jain R, et al. Impact of a national specialty e-consultation implementation project on access. Am J Manag Care 2015 Dec 01;21(12):e648-e654 [FREE Full text] [Medline: 26760427]
63. Keely E, Traczyk L, Liddy C. Patients' Perspectives on Wait Times and the Referral-Consultation Process While Attending a Tertiary Diabetes and Endocrinology Centre: Is Econsultation an Acceptable Option? Can J Diabetes 2015 Aug;39(4):325-329. [doi: 10.1016/j.jcjd.2014.12.010] [Medline: 25797111]
64. Liddy C, Maranger J, Afkham A, Keely E. Ten steps to establishing an e-consultation service to improve access to specialist care. Telemed J E Health 2013 Dec;19(12):982-990 [FREE Full text] [doi: 10.1089/tmj.2013.0056] [Medline: 24073898]
65. Kuziemsky C, Gogia S, Househ M, Petersen C, Basu A. Balancing Health Information Exchange and Privacy Governance from a Patient-Centred Connected Health and Telehealth Perspective. Yearb Med Inform 2018 Aug;27(1):48-54 [FREE Full text] [doi: 10.1055/s-0038-1641195] [Medline: 29681043]
66. Zhang A, Lin X. Towards Secure and Privacy-Preserving Data Sharing in e-Health Systems via Consortium Blockchain. J Med Syst 2018 Jun 28;42(8):140. [doi: 10.1007/s10916-018-0995-5] [Medline: 29956061]
67. Williams A, Larocca R, Chang T, Trinh N, Fava M, Kvedar J, et al. Web-based depression screening and psychiatric consultation for college students: a feasibility and acceptability study. Int J Telemed Appl 2014;2014:580786 [FREE Full text] [doi: 10.1155/2014/580786] [Medline: 24799895]
68. Westra I, Niessen FB. Implementing Real-Time Video Consultation in Plastic Surgery. Aesthetic Plast Surg 2015 Oct;39(5):783-790. [doi: 10.1007/s00266-015-0526-4] [Medline: 26169952]
69. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
70. Nutbeam D, McGill B, Premkumar P. Improving health literacy in community populations: a review of progress. Health Promot Int 2018 Oct 01;33(5):901-911. [doi: 10.1093/heapro/dax015] [Medline: 28369557]
71. Brant HD, Atherton H, Bikker A, Porqueddu T, Salisbury C, McKinstry B, et al. Receptionists' role in new approaches to consultations in primary care: a focused ethnographic study. Br J Gen Pract 2018 Jul;68(672):e478-e486 [FREE Full text] [doi: 10.3399/bjgp18X697505] [Medline: 29866710]
72. Liu L, Fang J. Study On Potential Factors Of Patient Satisfaction: Based On Exploratory Factor Analysis. Patient Prefer Adherence 2019;13:1983-1994 [FREE Full text] [doi: 10.2147/PPA.S228073] [Medline: 31819380]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
TAM: technology acceptance model

Edited by G Eysenbach, R Kukafka; submitted 26.06.20; peer-reviewed by J Li, MDG Pimentel, T Muto, M Adly, A Adly, A Adly; comments to author 11.08.20; revised version received 24.09.20; accepted 17.12.20; published 20.01.21.

Please cite as:
Qi M, Cui J, Li X, Han Y
Perceived Factors Influencing the Public Intention to Use E-Consultation: Analysis of Web-Based Survey Data
J Med Internet Res 2021;23(1):e21834
URL: http://www.jmir.org/2021/1/e21834/
doi:10.2196/21834
PMID:33470934

Review

Spread, Scale-up, and Sustainability of Video Consulting in Health Care: Systematic Review and Synthesis Guided by the NASSS Framework

Hannah M James[1], BKI; Chrysanthi Papoutsi[2], DPhil; Joseph Wherton[2], PhD; Trisha Greenhalgh[2], FMedSci; Sara E Shaw[2], PhD

[1]Department of Knowledge Integration, University of Waterloo, Waterloo, ON, Canada
[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Sara E Shaw, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Oxford,
United Kingdom
Phone: 44 1865617873
Email: sara.shaw@phc.ox.ac.uk

Abstract

Background: COVID-19 has thrust video consulting into the limelight, as health care practitioners worldwide shift to delivering care remotely. Evidence suggests that video consulting is acceptable, safe, and effective in selected conditions and settings. However, research to date has mostly focused on initial adoption, with limited consideration of how video consulting can be mainstreamed and sustained.

Objective: This study sought to do the following: (1) review and synthesize reported opportunities, challenges, and lessons learned in the scale-up, spread, and sustainability of video consultations, and (2) identify transferable insights that can inform policy and practice.

Methods: We identified papers through systematic searches in PubMed, CINAHL, and Web of Science. Included articles reported on synchronous, video-based consultations that had spread to more than one setting beyond an initial pilot or feasibility stage, and were published since 2010. We used the Nonadoption, Abandonment, and challenges to the Scale-up, Spread, and Sustainability (NASSS) framework to synthesize findings relating to 7 domains: an understanding of the health condition(s) for which video consultations were being used, the material properties of the technological platform and relevant peripherals, the value proposition for patients and developers, the role of the adopter system, organizational factors, wider macro-level considerations, and emergence over time.

Results: We identified 13 papers describing 10 different video consultation services in 6 regions, covering the following: (1) video-to-home services, connecting providers directly to the patient; (2) hub-and-spoke models, connecting a provider at a central hub to a patient at a rural center; and (3) large-scale top-down evaluations scaled up or spread across a national health administration. Services covered rehabilitation, geriatrics, cancer surgery, diabetes, and mental health, as well as general specialist care and primary care. Potential enablers of spread and scale-up included embedded leadership and the presence of a telehealth champion, appropriate reimbursement mechanisms, user-friendly technology, pre-existing staff relationships, and adaptation (of technology and services) over time. Challenges tended to be related to service development, such as the absence of a long-term strategic plan, resistance to change, cost and reimbursement issues, and the technical experience of staff. There was limited articulation of the challenges to scale-up and spread of video consultations. This was combined with a lack of theorization, with papers tending to view spread and scale-up as the sum of multiple technical implementations, rather than theorizing the distinct processes required to achieve widespread adoption.

Conclusions: There remains a significant lack of evidence that can support the spread and scale-up of video consulting. Given the recent pace of change due to COVID-19, a more definitive evidence base is urgently needed to support global efforts and match enthusiasm for extending use.

(J Med Internet Res 2021;23(1):e23775) doi:10.2196/23775

KEYWORDS

delivery of health care; remote consultation; telemedicine; videoconferencing; spread and scale-up; sustainability; mobile phone; COVID-19; remote care; consultation; review

Introduction

There is global interest in video consultation services, including Skype, Teams, FaceTime, and other Voice over Internet Protocol (VoIP) media, to facilitate synchronous patient-to-provider video communication in health care [1,2]. Studies have shown positive patient outcomes, reduced travel and costs, improved communication, decreased waiting times, and increased accessibility for patients [3]. Governments and health administrations view such technologies as a means of better managing demand and improving care [1]. The COVID-19 pandemic, and the requisite need for self-isolation and social distancing, has prompted rapid and widespread adoption of video consultations [4].

Evidence on the use of video consultations in health care is mixed. There is a rapidly growing literature on the feasibility, safety, acceptability, and effectiveness of video consultations across clinical areas [5] including diabetes [6,7], rehabilitation [8,9], mental health and addiction [10], cancer [11], palliative care [12,13], long-term care [14], geriatrics [15], postpartum support [16], and primary care [17]. Studies tend to be small scale and focus on initial adoption in a research context [2,5,18]; adopt a technology-centric approach (in which the technology is the primary focus, rather than the service or organization by which the technology is being used); and use trial methodology to study whether video consultation technology works or not. Despite calls for urgent action [19-22], this trend has continued into the COVID-19 pandemic. Little is currently known about how to successfully spread and scale up video consulting for sustained use across settings [23-25].

A small number of studies have explored the technological, contextual, and practical challenges to be overcome if video consulting is to become more widespread. One multilevel qualitative study, conducted in the English National Health Service (NHS) and undertaken by our team, examined the development, implementation, and use of video consultation services [2]. Focused on national-level policy, organizational-level implementation, and patient-clinician video consultations, the study identified a mismatch between the policy vision of video consultations replacing or supplementing a significant proportion of face-to-face care [26,27], and the substantial setup resources, ongoing human effort, and time needed to embed video consultations in routine care. Findings suggest that, even where there is significant policy impetus and demand [4], those implementing video consultation services face significant challenges in redesigning existing services and implementing new pathways.

Appreciation of the potential for longer term sustainability of this new service model is crucial in the context of the unfolding COVID-19 pandemic, which has brought a rapid need for alternatives to face-to-face contact and hence for the spread and scale-up of video consultations. We therefore conducted a systematic review of the opportunities and challenges to widespread implementation (what we refer to as "scale-up" or "spread") of video consultation services in health care, asking the following questions:

1. What theoretical frameworks have been used in this literature and for what purpose?
2. What opportunities and challenges have been identified in the literature on the spread, scale-up, and sustainability of video consultations?
3. What transferable insights can be useful for policy and practice? What questions remain unanswered?

Our main concern is to identify and evaluate challenges to the scale-up and spread of video consultations from the existing research literature and, from this, to inform the rollout and longer term sustainability of video consulting and the research agenda that can support it. Spread, scale-up, and sustainability are often used as interchangeable terms without a standard definition or adequate theorization (see Multimedia Appendix 1). In this review, we do not adopt a single definition of spread, scale-up, and sustainability, as we are interested in surfacing the different ways in which studies on video consultations have employed and operationalized these terms.

Methods

Information Sources and Search Strategy

In December 2018, we systematically searched 3 databases: PubMed, CINAHL via EBSCOhost, and Web of Science. The search was updated in March 2020. Our search strategies, developed with the help of a research librarian, used a mix of keywords, Medical Subject Headings (MeSH), and Major Headings (MH), as provided in Multimedia Appendix 2. We identified search terms bottom-up by examining titles, keywords, and frequently used phrases in relevant literature. For example, we derived terms such as "telemedicine" and "remote consultation" from literature on virtual health care and terms like "scalability" and "spread" from implementation science articles. We used broad telehealth and telemedicine terms to be able to address the variability in the terminology both in the scale-up and spread literature and in the way video consultation services have been described. The lead author reviewed PROSPERO (International Prospective Register of Systematic Reviews) prior to the study to identify similar reviews, which informed the search strategy and review focus. We referenced our initial search strategy against two published systematic reviews on telegenetics [28] and implementation science [29] to identify supplementary terms. Filters were applied to limit the results to published peer-reviewed articles. We focused on published literature and did not search grey literature. PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) was used to draft the protocol for this review (unpublished) and PRISMA guidelines were consulted throughout the review as a guide.

Eligibility Criteria

Inclusion criteria are summarized in Table 1. Included articles were peer-reviewed and reported on synchronous, video-supported consultations that had been scaled or spread to more than one setting beyond the initial pilot or feasibility stage (either within the same organization or to other organizations or geographic settings). The video consultation technology could be stand-alone, or part of a larger telehealth innovation (eg, a website). We focused on video consultation services connecting a patient to their medical provider(s) as opposed to those connecting providers to specialists. Articles exclusively reporting use, feasibility, acceptance, or pilot implementation with no evaluation of the implementation process were excluded. Video technologies like Skype have only recently come into use in health care [2,30], hence we restricted our search to articles published since 2010. Beyond these restrictions, we kept a broad interpretation of the patient-provider relationship, allowing video-to-home telehealth and hub-and-spoke consulting connecting a rural center to a specialist at a different hospital or institution. Our understanding of the quality of the articles was informed by using the Critical Appraisal Skills Programme (CASP) checklist for qualitative research and the Mixed Methods Appraisal Tool (MMAT) for mixed methods studies; however, we did not exclude articles on the basis of quality appraisal but took this into account in the interpretation of their findings (as reflected in the results section).

Table 1. Inclusion and exclusion criteria.

Criteria	Inclusion	Exclusion
Time period	2010 onwards	Before 2010
Language	English	Not English
Video consultation format	Synchronous and video-based conferencing OR digital health care technologies that include synchronous video conferencing.	Asynchronous, app-based, text-based, or website-based formats.
Actors	Patient and health care provider communication	Any other video consulting service
Context	Relevant technologies that have been scaled or spread to >1 setting in acute or primary care or where scale-up and spread is being actively pursued	Usability studies, feasibility trials, single-location implementation, pilot studies, or studies that had not been scaled or spread beyond one setting
Format	Peer-reviewed articles	Anything else (eg, conference proceedings, books, workshop papers)

Screening

Search results were imported into Zotero. Duplicates were removed, while retaining information from each search and preventing false duplicates from being merged. Each record was first screened by title and abstract (initially by HJ, then verified by JW and SS). Remaining articles were screened by reading the full texts (HJ screening 100%, SS a 15% sample), with any disputes resolved through discussion and consensus with JW. At each stage, articles were eliminated if they did not meet the inclusion criteria (Table 1). Full-text screening was iterative and required slight refinement of our inclusion criteria. We clarified that the articles must place sufficient emphasis on video consulting (rather than mention it in passing) and that different terms could be used to describe processes relevant to widespread implementation, scale-up, and spread.

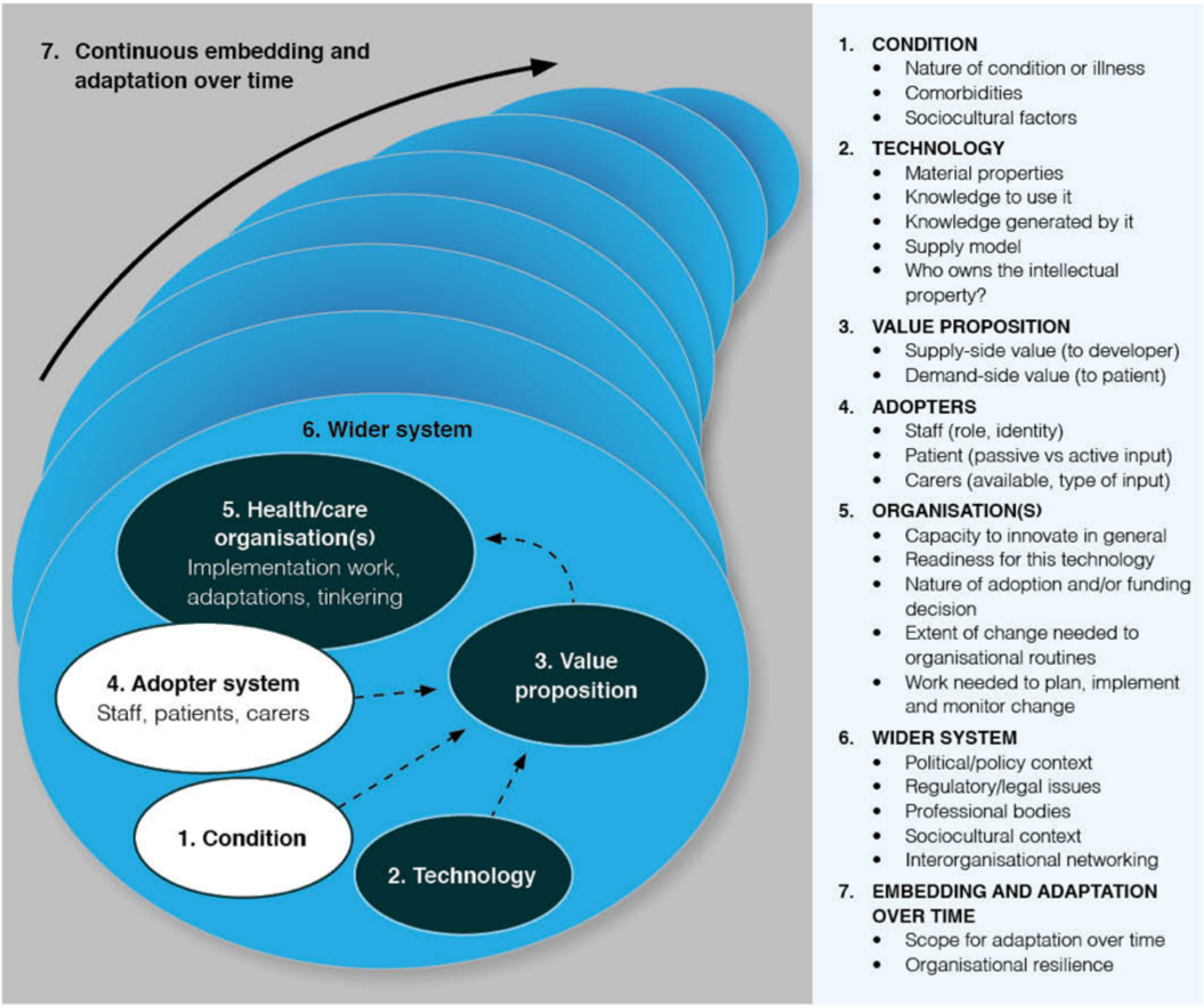
Data Extraction and Synthesis

We extracted data into a Microsoft Excel (Microsoft Corp) spreadsheet. At a descriptive level, we extracted data on study location, setting, clinical focus, definition of telehealth, service model and type of technology, and research design (Multimedia Appendix 3). We extracted data on how each of the papers conceptualized opportunities and challenges for the spread, scale-up, and sustainability of virtual consultations. We also extracted data about the predominant theoretical framework adopted in each study, and connected these with the three theoretical lenses that typically characterize studies of the spread and scale-up of innovations: (1) implementation science (ie, the systematic and structured application of improvement techniques and frameworks), (2) complexity theory (ie, paying attention to unpredictability and interdependencies in complex systems), and (3) social science approaches (ie, emphasizing social, human, and material elements in large-scale change efforts) [31].

We worked inductively to surface the challenges to and opportunities for the spread and scale-up of video consulting across studies. Sensitized by the Nonadoption, Abandonment, and challenges to the Scale-up, Spread, and Sustainability (NASSS) of health and care technologies framework [23], we then worked deductively to ensure we had not missed any of the 7 domains identified as relevant to the widespread adoption and nonadoption of innovations (ie, an understanding of the health condition[s] for which video consultations are being used, the material properties of the technological platform itself, the value proposition for patients and developers, the role of the adopter system, organizational factors, wider macro-level considerations, and emergence over time; see Figure 1).

We piloted and refined this process on one paper and then extended the process across the rest of our data set. Using content analysis, we compiled a descriptive overview of the opportunities and challenges identified in the papers, including frequency distributions. We then applied a more analytical lens to synthesize and interpret our findings using the NASSS framework [23].

Figure 1. The NASSS framework for Nonadoption, Abandonment, and challenges to Spread, Scale-up, and Sustainability of health and care technologies.



Results

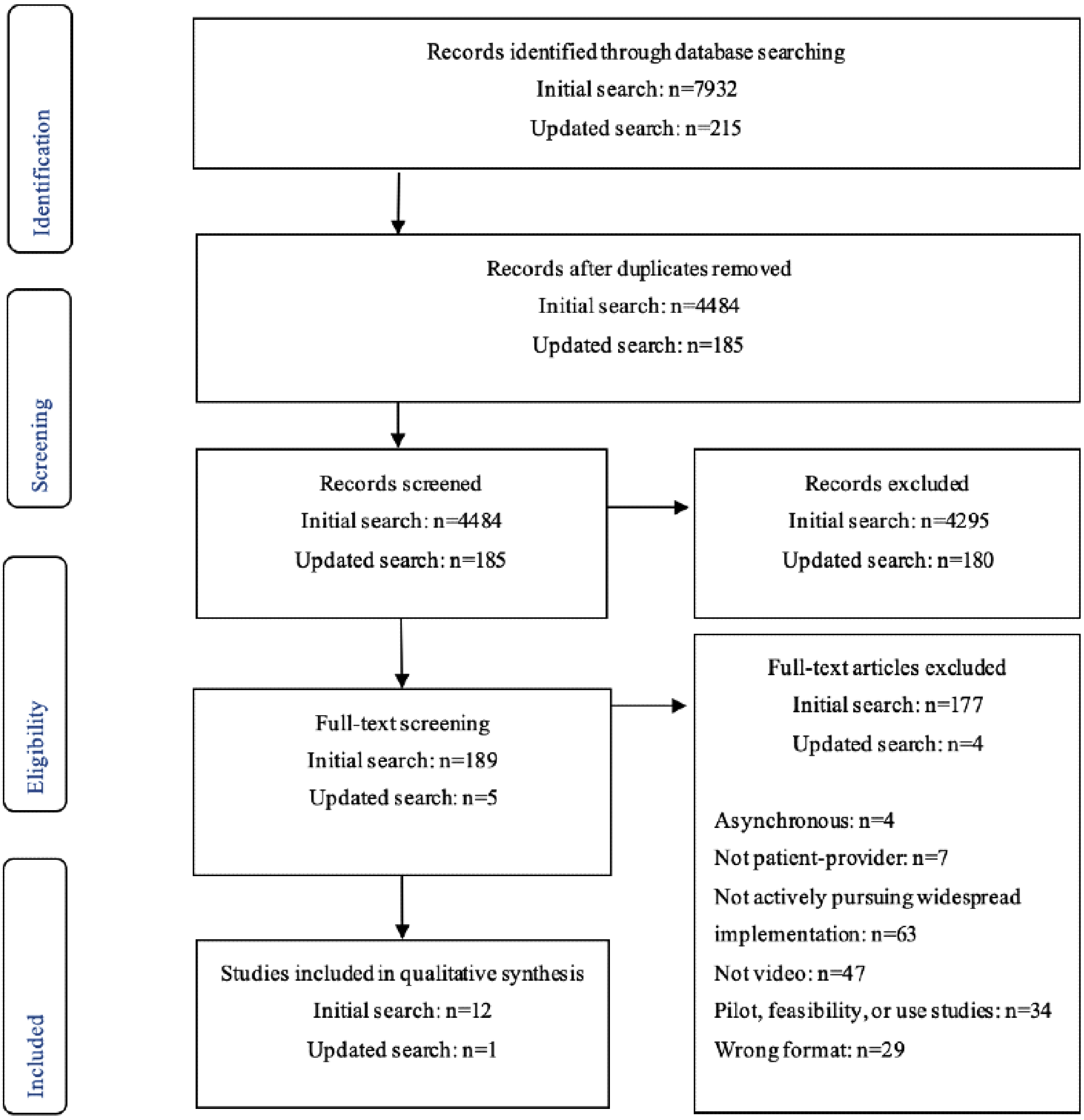
Description of Papers and Overview of Findings

Initial searches identified 4484 unique articles published between January 2010 and December 2018, with 193 articles identified for full-text screening and 12 initially included in the final review. An updated search in March 2020 returned 185 new articles for screening, of which 5 were identified for full-text screening. One new relevant article was identified, bringing the total records included in the review to 13 (Figure 2).

Diverse research designs were employed in the included studies, including interpretive case studies [32], structured or semistructured qualitative interviews [33-35], mixed methods combining qualitative and quantitative data [2,36-38], action research and deliberative methods [39], prospective implementation studies including quantitative activity and performance indicators [40-42], and retrospective case analysis of systemwide use of video consultations [43] (Multimedia Appendix 3). A review of the included studies using the CASP and MMAT appraisal tools indicated that they generally fulfilled relevant quality standards, although in a few articles it was not clear how data collection or analysis methods were applied, sample sizes were small, or there was inadequate information on sampling frame; furthermore, they remained at a highly descriptive or even purely illustrative level of analysis.

There was ambiguity in the use and variations of the term telehealth, with three studies giving no definition and others defining telehealth variably depending on study focus (Multimedia Appendix 3). The terms telehealth, telemedicine, and telemental health were common and often used interchangeably: 11 articles (85%) used the terms “video telehealth,” “telehealth,” “telemedicine,” or “telecare,” while two used the term “video consultations” [2,38].

Figure 2. PRISMA flow diagram depicting search and screening processes. PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis.



Papers described 10 different video consultation services in 6 regions: Australia [39], Cabo Verde [41], England [2,38], Nepal [32], Norway [36], and the United States (Multimedia Appendix 3). Studies described 3 types of video consultation services: (1) video-to-home services [2,34,38,39,42] that connected providers directly to a patient who received a video consultation on their own device; (2) hub-and-spoke models [32,33,35,37,40,41] that connected a provider at a central hub to a patient at a rural spoke center who called in with the assistance of a provider at that organization; and (3) large-scale top-down evaluations scaled up or spread across a single country [36] or a nationwide health administration [43]. Services covered various clinical specializations including rehabilitation [35,39], geriatrics [39], cancer surgery [2,38], diabetes [2,38], and mental health [34,37,40,42], in addition to general specialist care and primary care [32,33,36,41,43]. All studies presented empirical findings relating to services that had (to varying degrees) undergone, were undergoing, or were about to undergo scale-up or/and spread.

The different approaches and frameworks used in the papers reviewed reflect different logics of change. This has implications for the way spread, scale-up, and sustainability are operationalized, studied, and conveyed. Many of the articles included in the review emphasized a logic of change underpinned by implementation science (Multimedia Appendix 3). Bauer et al [37] used the Reach, Efficacy, Adoption, Implementation, and Maintenance (RE-AIM) framework in the quantitative arm of their evaluation on telehealth for bipolar disorder in the Veterans Affairs (VA) health care system to quantitatively assess the extent and patterns of implementation and sustainability. They also drew on the integrated Promoting Action of Research Implementation in Health Services (iPARIHS) framework to analyze qualitative data on challenges and opportunities for program implementation and sustainability. The Consolidated Framework for Implementation Research (CFIR; not designed specifically to look at implementation of software) was used in another implementation science–focused study to identify analytical constructs for collecting and analyzing provider perspectives on video telehealth in mental health services for US veterans [34,55]. Wade et al [39] used grounded theory to construct a process model of change management in large-scale home telehealth in South Australia, highlighting leadership support as a key facilitator. When reflecting on the transferability of their findings, this paper also makes reference to complex change and systems theory [39].

The above studies primarily report the use of frameworks as part of evaluating, rather than guiding, spread and scale-up efforts. Another two articles employed frameworks to support spread and scale-up efforts, although they did not draw on relevant approaches such as the Going to Full Scale framework, 3S infrastructure, or the Dynamic Sustainability framework [58-60]. In a prospective analysis in Cabo Verde, the “initiate-build-operate-transfer” approach was described as the basis for delivering a countrywide telemedicine network. This seemed akin to a phased implementation framework, although it was supplemented by a range of additional measures to address sustainability factors as reported in relevant literature, including careful training and capacity development [41]. In a second prospective study, the PARIHS framework, including use of external facilitation, was employed as a systematic approach to guide extensive implementation of psychotherapy for posttraumatic stress disorder (PTSD) by the US Department of Veterans Affairs [40].

Complexity was mentioned in a few articles, although not as a key organizing framework. Alami et al [36] analyzed data from their mixed methods study on national telemedicine implementation in Norway without employing a conceptual framework; however, their discussion emphasizes complexity, adaptive capability, and participatory approaches. Darkins [43] drew on diffusion of innovations theory to organize his analysis of telehealth expansion in the US Veterans Health Administration over a 10-year period, referring as well to the “complex adaptive environment” and “systems approach” taken to achieve spread.

Social science theory was also employed, although rarely in a highly theoretical mode. In their study on telemental health services for rural American Indian communities, Brooks et al [33] used diffusion of innovations theory to retrospectively examine factors that influenced widespread adoption. Martinez et al [35] applied a sociotechnical perspective in their study of health providers’ perspectives on video telehealth for US veterans with spinal cord injuries. This led to identification of social and technical factors that influenced telehealth implementation across care facilities. In their discussion, they also reflected on the interdependencies and relationships between the different sociotechnical aspects of the system, viewing health care teams as complex adaptive systems. Bhatta et al [32] presented a descriptive overview of opportunities and challenges for telemedicine in Nepal and included some, albeit limited, discussion of their findings from an information infrastructures perspective to highlight the importance of installed base. Finally, the VOCAL study took an explicit social science approach by theorizing video consultations using technology-enhanced Strong Structuration Theory, which assumes a dynamic and reciprocal link between the social environment, human interpretations and actions, and technologies [2,38].

Reported Opportunities and Challenges

A total of 38 opportunities to scale-up, spread, and sustainability and 47 unique challenges were reported across the 13 articles. The most common opportunities included the availability of clinical and/or nonclinical telehealth champions or coordinators (n=8), provider acceptance (n=4), absence of billing or licensure restrictions (n=3), adequate funding (n=4), and strong interorganizational communications (n=4). The most common challenges were lack of technical telehealth-specific support for clinical staff (n=6), need for redevelopment of workflows and organizational routines (n=6), financial pressures (n=5), and lack of training (n=5).

Not all articles explicitly reported opportunities and challenges: 10 (77%) reported both challenges and opportunities [2,33-35,37-41,43], 2 (15%) reported challenges exclusively [32,36] and 1 (8%) article focused more on opportunities and how these could be translated to other settings [42].

Challenges to Scale-up, Spread, and Sustainability

We synthesized the data extracted from the 13 articles according to the 7 domains of the NASSS framework [23] (Figure 1).

Domain 1: The Condition

This domain encompasses the clinical and sociocultural aspects of the health condition and associated comorbidities, acknowledging that not all individuals with the same condition would benefit equally from health technologies [23]. The complexity of the condition plays a role in determining the suitability of patients for video consultations and hence influences the potential for scale-up, spread, and sustainability [23]. In total, 7 of the studies reported the conditions for which video consultations were used, including dermatology, diabetes, antenatal diabetes, postoperative cancer, spinal cord injuries, bipolar disorder, PTSD, and other mental health conditions. Of these studies, 6 paid limited attention to how clinical characteristics played a role in the successful spread of video consulting services [32-35,37,42], although there was mention of provider concerns around using video consulting in specific conditions (eg, “for patients who are at high risk of suicide or who have psychotic symptoms”). This was mainly reported by participants with no previous experience of video consulting [34].

Opportunities for and challenges to the spread and scale-up of video consulting by type of clinical condition were reported in detail in only one study [2,38]. For example, video consulting in an antenatal diabetes clinic was abandoned given the involvement of multidisciplinary teams across departments, the duration and severity of the condition (short-term, high-risk), and the use of patient-held medical records that were unavailable to the clinician conducting the video consultation [2,38]. Preoperative cancer surgery was deemed too complex for video consulting given the necessity of a physical exam combined with there likely being no pre-existing relationship between the patient and clinical team, but postoperative follow-up within the same clinic was deemed more appropriate [2,38]. Other studies pointed to the difficulties in conducting physical examinations as constraining spread and scale-up. For example, an investigation of video consulting scale-up for spinal cord injury suggested that a “common concern was the perceived limits to evaluate physical symptoms over [video]” and the difficulty to assess clients’ complaints without being able to physically examine them [35]. One study noted that for patients with agoraphobia, the video medium removed the need to travel to the clinic and allowed patients to more easily receive mental health support [42].

Domain 2: The Technology

The technology domain encompasses the materials, data, knowledge, and supply features of the technology [23]. This domain focuses not only on the physical technology and knowledge required to use it, but also on how the technology shapes (and becomes shaped by) the potential for scale-up and spread [23]. The 13 papers referred to various video consultation programs but focused on their implementation process rather than providing details of the video conferencing technologies used. Some descriptions remained vague; for example, one article referred to a “simple technology that is only used for consultation purposes” [32]. Others provided a brief overview of changes in technological solutions, from early telehealth programs that used “commercial off-the-shelf videoconferencing systems, ones that had been developed for administrative, not clinical purposes” to latter stages using more sophisticated technological tools, such as teleretinal imaging [43]. Companion devices were also mentioned as providing opportunities for enhanced video consulting, including “an e-stethoscope, vital signs monitor, and dermatology camera, but also the e-electrocardiograph and ultrasound probe” [41]. Little emphasis was placed on describing the material features of the technologies; however, 6 studies reported technical challenges restricting spread, including the following: lack of reliability of the video conferencing technology including audio or video interruption/failure [2,32,34,35,38], inadequate maintenance [32], and unreliable internet accessibility or bandwidth [32,35,41].

Brooks et al [33] also reported on challenges setting up the infrastructure to be able to accommodate video consulting: “respondents noted many challenges in the clinic implementation process. Among these were … setting up the telehealth backbone.” In Norway, infrastructural differences between the different regions have challenged the spread of telemedicine and meant that the national eHealth strategy could not be implemented consistently [36]. Elsewhere, unreliable or inconsistent national internet capability and service was reported as a challenge to implementation [32,41]. For example, in Nepal, “due to the irregular supply of electricity and slow internet service, it is difficult … during video conference consultations due to frequent disconnection, blurry images, and unclear sound” [32]. Additionally, widespread implementation requires patients to have access to the appropriate “infrastructure” to participate in a video consultation (computer, webcam, data allowance) [40].

Across all papers reviewed, available technologies for video consultations posed challenges for establishing remote models of care as well as spread and scale-up. Authors reported that support and high-quality training were necessary to achieve widespread growth [2,32-43]. This was illustrated strongly in a large-scale implementation in Nepal: “Another challenge was related with lack of competence and training among the personnel involved in the rural-telemedicine program” [32]. Even when specialist information technology (IT) staff were available, a lack of service and procurement standards hindered spread across the system, as in the example of video consulting in the Veterans Health Administration: “telehealth projects were developed using bespoke interfaces at an individual medical center, ones that were not replicable across the system. Both Veterans Health Administration (VHA) IT and biomedical engineering supported telehealth but without consistent standards for equipment purchase, installation, service, warranty, and help-desk arrangements” [43].

Domain 3: The Value Proposition

The value proposition domain is concerned with whether or not the technology is worth developing or introducing for clinicians, patients, and suppliers [23]. It can influence upstream supply of the technology and uptake and desirability on the demand side, impacting scale-up and spread [23]. The majority of papers (n=10) discussed the value proposition for video consulting and how this was articulated by different program stakeholders. Value proposition was primarily understood in terms of clinical stakeholder perception of patient demand and interest [39,42], perceived clinical need/utility [2,32,33,35,38,41,43], and supplier benefit [2,38]. For example, in an evaluation of video consulting for individuals with bipolar disorder, growth was enabled by technology that was “successfully filling a need perceived by providers” [37]. In another study, Lindsay et al [42] reported that the “natural disaster Hurricane Harvey offered unique motivation for previously reluctant providers to use VTH delivery to connect with their patients during the crisis and beyond,” not only emphasizing the value proposition changing mid-implementation, but also enabling rapid scale-up in ways not previously imaginable.

A weak or poorly articulated value proposition posed challenges for scale-up and spread (n=2). For example, Martinez et al [35] reported that “some providers encountered initial hesitance from individuals with Spinal Cord Injuries or Disorder which they mostly attributed to patients’ uncertain feelings about the new technology.” Interian et al [34] also described how a video consulting program aimed at improving access to mental health care did not spread successfully in an urban environment compared to other settings, because providers did not perceive a local need for the technology: “A second issue involving provider buy-in pertained to the perceived local need for implementation.”

The upstream supply of technology (ie, the relationship between the organization and the supplier or developer) was reported as a facilitator by Greenhalgh et al [38], who explained that scale-up in a Skype-based service was facilitated partially by “clear benefit for both the technology supplier and the patient.” Contrastingly, Bhatta et al [32] reported how uncertain supplier relationships and funding threatened the sustainability and scale-up of their program in Nepal.

Domain 4: The Adopter System

The adopter system considers the staff, patients, and caregivers and their potential and desire to adopt and to continue to use a technology [23]. Staff concerns with video technologies were reported in all papers (n=13) and included the following: lack of provider buy-in and resistance to change [2,33,36,38,39,42]; lack of adequate training and telehealth staff to support sustainable operation of the video consulting programs [32,34,35,37,41], and the mobility of human resources within and between systems including “unplanned transfer of healthcare workers” [32]; the “high turnover of clinicians” [36], and availability of staff, space, and equipment [37]. Providers were slow or reluctant to buy in because they did not trust the technology or consider it clinically applicable [33,34,42]; did not have experience with the technology [34,40]; did not have adequate time to devote to the new technology [2,38]; did not have enough evidence to support video consulting use, resulting in resistance to changes in care models, as they found existing care models to adequately fulfill patient needs [39,40]; or lacked training and technological literacy or ongoing support [32,35].

Spread and scale-up were enabled by an adopter system with engaged and committed staff who were given protected time by the organization to conduct video consultations, as mentioned by Interian et al [34]: “having a team of providers who were detailed solely to provide mental health services by using telehealth technology and had protected time to do so… was rated as having a strong positive impact on V2H [video-to-home telehealth] implementation.” Engaged leadership staff were also reported to be pivotal for promoting uptake and creating conditions for sustained and widespread provider uptake [42].

Challenges to and opportunities for widespread implementation for patients and caregivers were reported in 6 articles [2,32-34,38,42]. Reported patient and caregiver challenges to and opportunities for widespread implementation included patient trust and acceptance of video consulting [2,33,38], the patient’s ability to transport themselves to a “hub” site (for hub-and-spoke video consulting [33]), and patients requiring appropriate technology to participate in video consultations [34]. In Nepal, challenges for patients were numerous and included cultural barriers (eg, patients’ feelings of inferiority to health care workers; women and older patients not speaking for themselves; lack of confidence) and literacy concerns, which were reported to make “it difficult and time consuming to use [video consults] for such patients” [32]. When patients were in front of a screen, it was reported by staff that patients felt uncomfortable participating in the consultation. For mental health care in the United States, patients reported liking the convenience and privacy of video consulting; however, these responses were collected from patients with repeated prior use of video consulting services, possibly excluding important challenges of patients with little or no experience with video consultations [42].

Despite the abundance of challenges to and opportunities for widespread implementation reported from the provider perspective, there was little in-depth consideration of the impact of patient and caregiver adoption on widespread implementation and how the work and acceptance of these groups may have influenced widespread and sustainable implementation. Of the articles that did report patient or caregiver challenges [2,32-34,38,42], these were primarily clinician perceptions, rather than direct reports from patients or caregivers.

Domain 5: The Organization

The organization domain considers the capacity of the organization to innovate, its readiness for change, the nature of the funding decision, the extent of change in routines, and the work needed to implement change in the organization as it relates to the new technology [23]. These elements are crucial in scale-up, spread, and sustainability because they address how the organization might respond to emerging challenges and how the innovation coevolves with organizational structures and processes [23]. Challenges reported in 5 articles included the labor-intensive process of scheduling video consultation appointments within and across organizations and the need to redesign organizational processes around the technology [2,36-38,40]. As Bauer et al [37] reported: “The scheduling process was noted to be labor-intensive, requiring a three-way match among the veteran’s schedule, the consultant’s schedule, and telehealth room availability.” Other challenges included how sites with existing video consulting programs and pre-existing routines may be hesitant or slower to implement new technologies and redesign workflows around a new technology [40] and how embedding video consultations in a clinic “involved significant reworking of … processes in ways that took account of the ‘virtual’ presence of the patient” [38]. In some cases, redesign of organizational processes was prohibitive to implementation. For instance, lack of widespread video consulting implementation in Norway was at least partially attributed to the lack of organizational preparedness required to “integrate changes and initiate restructuring” [36].

Complexity in the organizational domain was managed more effectively when telehealth champions were available, as reported in 11 of the reviewed articles [2,33-37,39-43]. In the words of one author, telehealth champions were “vital to [the] growth of telehealth” [43]. Telehealth champions were typically clinical staff [2,34,36,41-43]. Other roles included health care coordinators or IT staff [33,35,40,43].

Domain 6: The Wider Context

The wider context domain considers the political, regulatory, professional, and sociocultural aspects of video consulting implementation [23]. These factors shape the context in which the technology is implemented and influence the potential for growth and sustainability of video consulting services. At least one challenge or opportunity related to political, regulatory, cultural, contextual, or financial factors was reported in all studies (n=13) [2,32-43]. Political challenges included government instability, which affected related policy development, as a study in Nepal explained: “Due to the unstable political situation and frequent change of government, the policy related with the health care delivery system is fragile” [32]. By contrast, in the island country of Cabo Verde in Western Africa, there was a high-level political mandate for video consulting and a favorable collaboration environment with clear responsibilities between stakeholders, including international nongovernmental organizations. This facilitated the nationwide scale-up and spread of video consulting: "…the collaboration among the … government, the donor, and the implementing agency is superb, with clear terms of references for each of the players" [41].

Even when the political environment was more stable, papers reported that national strategies did not always emphasize the potential for scale-up and spread against other priorities. Take the following report from an evaluation in Norway, which suggests that national approaches "were mainly focused on messaging services and electronic exchange, while less attention was paid towards telemedicine services, especially videoconferencing" [36]. There was a gap in the papers reviewed, on the relevance of interorganizational networking for supporting and sustaining the spread and scale-up of video consulting [24]. Brooks et al [33] highlighted the importance of information (eg, background information about services, clinical protocols) from other organizations using telehealth in promoting take-up and spread. Two other papers reinforced this [2,43], referring to, for instance, "telehealth communities," "collaborations across clinics and staff," and "collaborations with…government partners" not only as important catalysts for change, but also as critical to the ongoing evolution and spread of services. None of the papers mentioned specific interorganizational initiatives (eg, quality improvement collaboratives). This was a surprising silence given the recognized importance of interorganizational networking for the spread of innovation [44,55,61].

Absence of or ambiguity in reimbursement was another common contextual barrier that influenced wider spread of video consultations [36]. In a UK study, the authors reported that, although lack of clarity on remuneration was often raised as a key barrier by implementing teams, published policy documents rarely explained how "reimbursement for virtual consultations would be implemented" [38]. In contrast, centralized funding for video consulting in the Veterans Health Administration in the United States meant that the "availability of national support infrastructure and the absence of billing for services were distinct advantages" [37].

Domain 7: Interactions Between Domains and Adaptation Over Time

The final domain, embedding and adaptation over time, focuses on the scope for adaptation and the resilience of the organization in the face of implementation and potential spread and scale-up [23]. One study explicitly recognized interactions between domains [2,38] by looking across three levels of data relating to spread and scale-up of video consulting: micro (individual users), meso (organizational processes and systems), and macro (national policy and wider context) [2,38]. They reported how challenges such as the condition interact with the organization, determining the ability to scale up video consulting services. For example, in a scale-up of 4 video consulting services, the challenges and opportunities for each service were connected across the domains; simply assessing a patient as having a theoretically appropriate condition for video consulting (micro) and having a supportive national policy (macro) did not guarantee success. Their specific interacting challenges included financial challenges, organizational structure, technical challenges, and the existing structure of treatment [2,38].

Factors impacting resilience, "the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances so that it can sustain required operations, even after a major mishap or in the presence of continuous stress" [62], and adaptation over time were reported in three articles [32,42,43]. Notably, the absence of sustainable financial models was reported to threaten sustainability [32]. However, in a comprehensive description of the 20-year evolution and adaptation of telehealth within the VHA in the United States, Darkins [43] illustrates a dynamic transition from a fragmented health care system with many short-term pilot telehealth implementations that slowly scaled up and increased the scope of video consulting services and the associated infrastructure [43]. Lindsay et al [42] provided detailed resolution of adaptation over time for one health center within the VHA. In this implementation project, resiliency and sustainability were enabled by introducing video consulting into general practice rather than training specific providers and using a flexible implementation approach that can be modified for different health system contexts. Furthermore, the authors emphasized how measuring and assessing outcomes of implementation and sustainability then sharing those results with stakeholders "increases motivation and momentum for practice change and enables site to respond to challenges in real time" [42].

Discussion

Principal Results

This systematic review contributes valuable insights about the potential for scale-up, spread, and sustainability of video consulting and a novel interpretation through the application of the NASSS framework. A key finding is the paucity of evidence thoughtfully articulating challenges to the scale-up and spread of video consulting, which sits uncomfortably alongside the current global enthusiasm for expanding the use of video consulting services. The review also reveals an absence of concrete operationalization and theorization of scale-up and spread, exemplified by the absence of analysis and concrete definitions and terminology, as well as a lack of appreciation of complexity (particularly, but not only, in relation to the clinical condition and organizational implementation). The articles view spread and scale-up as the sum of multiple implementations, rather than theorizing the distinct processes required to achieve widespread adoption.

Findings from the review pointed to potential enablers of the spread and scale-up of video consulting services, notably regarding the influence of the presence of a telehealth champion, especially during initial spread [63], as well as leadership at multiple levels, appropriate reimbursement mechanisms, user-friendly technology, pre-existing relationships between staff [64], and adaptation (of technology and services) over time [65]. Findings also raised a number of challenges including the following: technically challenged staff, resistance to change, cost, reimbursement, and patient characteristics [24], as well as project management, patient recruitment, leadership involvement, funding, absence of a long-term strategic plan, resistance to change, workflow changes, lack of resources, and liability [64-66]. These challenges are consistent with previous research concerning small-scale implementation and diffusion of video consulting services and the scale-up and spread of other eHealth interventions [67,68]. They also indicate that the same things that interfere with implementation at one site [eg, 48] are also seen as interfering with spread and scale-up [5,7,17,24], although this may be an artifact of how spread is commonly studied as the sum of multiple implementations.

There are gaps in the existing evidence on the spread and scale-up of video consulting. This was particularly the case in relation to interorganizational networking, which is critical in enabling and sustaining innovation [24,30,64], and yet is not explored in-depth. There is also limited data on the material properties and design of different video consulting platforms, and on the value proposition in relation to spread and scale-up when a technology is free (eg, part of a research study) in the place where it was initially developed or implemented, but potentially costly in other settings.

As patients, health care organizations, and nations continue to look toward video consultations, it is essential to continue to theorize this domain. The absence of consistent theorization was demonstrated in this review by the absence of consideration of how challenges and opportunities might interact with each other and influence the process. One proposal to address this oversight is the theorization of scale-up and spread as social processes [69], prioritizing the interactions between actors and context to better inform scale-up or spread. Our review specifically draws on the NASSS framework to contribute a different way of viewing the spread and scale-up of video consultations, understanding opportunities and challenges as emergent, in constant tension, and inherently social. In conditions of complexity, spread and scale-up efforts may be further supported through the facilitation of interdependencies and relational aspects of change, processes that allow sense-making and experimentation, and scope for local adaptations and self-organization [31,70].

The COVID-19 pandemic is producing what is essentially a “natural experiment” as alternatives to traditional face-to-face consultations become a necessity to prevent viral transmission [4]. Interestingly, albeit much more localized, Hurricane Harvey was reported by Lindsay et al [42] to be an opportunity for provider buy-in for video-mediated mental health services. This suggests that the wider context of the COVID-19 pandemic might similarly enable widespread adoption of video consulting services.

Limitations

This review is limited by likely bias within the articles, bias inherent in the reviewing process, and the lack of theorization in the included studies. Publication bias is a long-standing and recognized phenomenon in health research [71] and likely impacts the broadly positive results across studies included in this review. Given the lack of theorization in the included studies, it is hard to know how broadly the factors that may have acted as challenges or opportunities were considered in the evaluations. The consequence of this is that we cannot claim that this evidence provides a conclusive list of the challenges to scale-up and spread generally, nor even specifically to the studies included in this review.

The strengths of this review include the comprehensive searching methods used to identify a breadth of published research (albeit limited in numbers and scope). The analysis and synthesis of the listed challenges has also been framed by a comprehensive theoretical framework.

Comparison With Prior Work

Our review contributes an important synthesis of knowledge about the spread, scale-up, and sustainability of video consulting. Our analysis, guided by the NASSS framework, has enabled us to examine a diverse body of evidence on video consulting services that have scaled or spread to more than one setting. This addresses a fundamental gap in the literature, with prior studies typically focusing on individual technologies or services, rather than organizational implementation and spread across settings [5].

To date, limited attention has been paid to the spread and scale-up of video consulting. Previous systematic reviews have tended to isolate areas of specific interest, focusing on selected aspects of video consulting, such as patient satisfaction [3,72], feasibility [5], clinical effectiveness and cost-effectiveness [73], and specialized clinical areas (eg, cardiology [74], diabetes [75], mental health and addiction [10], teledermatology [76]). Additionally, reviews have focused on telemedicine or eHealth in general without focusing on video consulting [24,77], with which our findings are consistent. Prior reviews have also focused exclusively on the technical aspects of video consulting technology [7]; considered technologies at any stage of implementation, rather than technologies with widespread adoption [24]; or restricted the types of articles included in the review to randomized control trials [78]. Studies have drawn on specific implementation frameworks to support and theorize spread and scale-up of video consulting, with a mechanistic logic of change being dominant. Other frameworks do exist that might support spread and scale-up (see [24] for an overview) but they have yet to be taken up in the context of expanding video consulting.

The studies we reviewed are broadly positive about video consulting. However, the small sample sizes, select nature of samples, and high losses to follow-up call into question any unqualified conclusion that video consulting is “effective.” The trials that have been undertaken on video consulting have provided few or no data on the organizational complexities of implementing a technology-based service, and do not address the question of how video consultation services become embedded in real-world settings.

To date, there have been only a handful of rigorous and theoretically grounded qualitative or mixed methods studies that explore the emergence of video consultation services naturalistically. Such studies have yet to systematically study spread and scale-up beyond the initial implementation.

Conclusions

This systematic review identified opportunities and challenges for the scale-up and spread of video consulting. The application of the NASSS framework surfaced complexity in a number of domains, notably characterizing the importance of organizational support and the wider system in the scale-up and spread of video consulting services. Many articles viewed spread and scale-up purely as the sum of multiple implementations, rather than explicating and theorizing the distinct (social, interorganizational, and policy-related) processes required to achieve widespread adoption.

Given the recent impetus to implement video consulting services at pace and scale due to the COVID-19 pandemic, a more definitive evidence base is urgently needed to support global efforts, and match policy enthusiasm for the widespread use of these technologies. We strongly encourage decision makers and researchers to embrace relevant theoretical lenses that can aid spread and scale-up and ensure the future sustainability of what looks set to be a significant part of future health care delivery.

Acknowledgments

This study formed the basis of an unfunded bachelor's dissertation project. HJ devised the study with SS; led the process of identifying and reviewing papers, with support from SS and JW; and wrote an initial first draft of the paper for her dissertation. Following the start of the COVID-19 pandemic, CP, SS, and HJ theorized spread, scale-up, and sustainability; updated the search; and further built up the analysis with input from TG and JW. All authors contributed to drafts and approved the final version of the paper. Our thanks go to library services at the University of Waterloo for support with searching.

Input on the review from TG, SS, JW, and CP is supported by a Wellcome Trust Senior Investigator Award to TG (grant reference WT104830MA) and funding from the Health Foundation for research on "Video consulting during and beyond the COVID-19 pandemic: implications for scale up and sustainability" (grant reference 2133488). SS and TG are partly funded by the NIHR Biomedical Research Centre (BRC), Oxford, United Kingdom (grant reference number NIHR-BRC-1215-20008). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Spread, scale-up, and sustainability definitions: an overview.
[DOCX File , 26 KB - jmir_v23i1e23775_app1.docx]

Multimedia Appendix 2
Search terms used.
[DOCX File , 18 KB - jmir_v23i1e23775_app2.docx]

Multimedia Appendix 3
Summary of papers included in the review.
[DOCX File , 21 KB - jmir_v23i1e23775_app3.docx]

References

1. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
2. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Southampton (UK): NIHR Journals Library; 2018.
3. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
4. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
5. Armfield NR, Bradford M, Bradford NK. The clinical use of Skype--For which patients, with which problems and in which settings? A snapshot review of the literature. Int J Med Inform 2015 Oct;84(10):737-742. [doi: 10.1016/j.ijmedinf.2015.06.006] [Medline: 26183642]
6. McDonnell ME. Telemedicine in Complex Diabetes Management. Curr Diab Rep 2018 May 24;18(7):42. [doi: 10.1007/s11892-018-1015-3] [Medline: 29797292]
7. Fatehi F, Armfield NR, Dimitrijevic M, Gray LC. Technical aspects of clinical videoconferencing: a large scale review of the literature. J Telemed Telecare 2015 Feb 19;21(3):160-166. [doi: 10.1177/1357633x15571999]

8. Phillips VL, Temkin AJ, Vesmarovich SH, Burns R. A feasibility study of video-based home telecare for clients with spinal cord injuries. J Telemed Telecare 1998;4(4):219-223. [doi: 10.1258/1357633981932271] [Medline: 10505358]
9. Knox L, Dunning M, Davies C, Mills-Bennet R, Sion TW, Phipps K, et al. Safety, feasibility, and effectiveness of virtual pulmonary rehabilitation in the real world. Int J Chron Obstruct Pulmon Dis 2019;14:775-780 [FREE Full text] [doi: 10.2147/COPD.S193827] [Medline: 31040656]
10. Kruse CS, Lee K, Watson JB, Lobo LG, Stoppelmoor AG, Oyibo SE. Measures of Effectiveness, Efficiency, and Quality of Telemedicine in the Management of Alcohol Abuse, Addiction, and Rehabilitation: Systematic Review. J Med Internet Res 2020 Jan 31;22(1):e13252 [FREE Full text] [doi: 10.2196/13252] [Medline: 32012048]
11. Bensink ME, Armfield NR, Pinkerton R, Irving H, Hallahan AR, Theodoros DG, et al. Using videotelephony to support paediatric oncology-related palliative care in the home: from abandoned RCT to acceptability study. Palliat Med 2009 Apr;23(3):228-237. [doi: 10.1177/0269216308100251] [Medline: 19073783]
12. Funderskov KF, Raunkiær M, Danbjørg DB, Zwisler A, Munk L, Jess M, et al. Experiences With Video Consultations in Specialized Palliative Home-Care: Qualitative Study of Patient and Relative Perspectives. J Med Internet Res 2019 Mar 21;21(3):e10208 [FREE Full text] [doi: 10.2196/10208] [Medline: 30896436]
13. Jess M, Timm H, Dieperink KB. Video consultations in palliative care: A systematic integrative review. Palliat Med 2019 Sep;33(8):942-958. [doi: 10.1177/0269216319854938] [Medline: 31187688]
14. Wakefield BJ, Buresh KA, Flanagan JR, Kienzle MG. Interactive video specialty consultations in long-term care. J Am Geriatr Soc 2004 May;52(5):789-793. [doi: 10.1111/j.1532-5415.2004.52220.x] [Medline: 15086663]
15. Shulver W, Killington M, Morris C, Crotty M. 'Well, if the kids can do it, I can do it': older rehabilitation patients' experiences of telerehabilitation. Health Expect 2017 Feb;20(1):120-129 [FREE Full text] [doi: 10.1111/hex.12443] [Medline: 26890025]
16. Vincze L, Rollo ME, Hutchesson MJ, Callister R, Collins CE. VITAL change for mums: a feasibility study investigating tailored nutrition and exercise care delivered by video-consultations for women 3-12 months postpartum. J Hum Nutr Diet 2018 Jun;31(3):337-348. [doi: 10.1111/jhn.12549] [Medline: 29543356]
17. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
18. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized Telehealth in the Future: A Global Research Agenda. J Med Internet Res 2016 Mar 01;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]
19. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
20. Lonergan P, Washington III SL, Branagan L, Gleason N, Pruthi R, Carroll P, et al. Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
21. Neubeck L, Hansen T, Jaarsma T, Klompstra L, Gallagher R. Delivering healthcare remotely to cardiovascular patients during COVID-19 : A rapid review of the evidence. Eur J Cardiovasc Nurs 2020 Aug;19(6):486-494 [FREE Full text] [doi: 10.1177/1474515120924530] [Medline: 32380858]
22. Trethewey SP, Beck KJ, Symonds RF. Video consultations in UK primary care in response to the COVID-19 pandemic. Br J Gen Pract 2020 May;70(694):228-229. [doi: 10.3399/bjgp20X709505] [Medline: 32354807]
23. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
24. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
25. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013 Feb;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
26. NHS. The NHS Long Term Plan. London: NHS; 2019. URL: https://www.longtermplan.nhs.uk/ [accessed 2021-01-19]
27. Virtual Care Task Force. Virtual Care: Recommendations for Scaling up Virtual Medical Services. Ottawa, Canada: Canadian Medical Association; 2020 Feb. URL: https://www.cma.ca/sites/default/files/pdf/virtual-care/ReportoftheVirtualCareTaskForce.pdf [accessed 2021-01-19]
28. Hilgart JS, Hayward JA, Coles B, Iredale R. Telegenetics: a systematic review of telemedicine in genetics services. Genet Med 2012 Sep 12;14(9):765-776. [doi: 10.1038/gim.2012.40] [Medline: 22498847]
29. Ben Charif A, Zomahoun HTV, LeBlanc A, Langlois L, Wolfenden L, Yoong SL, et al. Effective strategies for scaling up evidence-based practices in primary care: a systematic review. Implement Sci 2017 Nov 22;12(1):139 [FREE Full text] [doi: 10.1186/s13012-017-0672-y] [Medline: 29166911]

30. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic Consultation in Primary Care Between Providers and Patients: Systematic Review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
31. Greenhalgh T, Papoutsi C. Spreading and scaling up innovation and improvement. BMJ 2019 May 10;365:l2068 [FREE Full text] [doi: 10.1136/bmj.l2068] [Medline: 31076440]
32. Bhatta R, Aryal K, Ellingsen G. Opportunities and Challenges of a Rural-telemedicine Program in Nepal. J Nepal Health Res Counc 2015;13(30):149-153. [Medline: 26744201]
33. Brooks E, Manson SM, Bair B, Dailey N, Shore JH. The diffusion of telehealth in rural American Indian communities: a retrospective survey of key stakeholders. Telemed J E Health 2012 Jan;18(1):60-66 [FREE Full text] [doi: 10.1089/tmj.2011.0076] [Medline: 22082106]
34. Interian A, King AR, St Hill LM, Robinson CH, Damschroder LJ. Evaluating the Implementation of Home-Based Videoconferencing for Providing Mental Health Services. Psychiatr Serv 2018 Jan 01;69(1):69-75. [doi: 10.1176/appi.ps.201700004] [Medline: 28859586]
35. Martinez RN, Hogan TP, Balbale S, Lones K, Goldstein B, Woo C, et al. Sociotechnical Perspective on Implementing Clinical Video Telehealth for Veterans with Spinal Cord Injuries and Disorders. Telemed J E Health 2017 Jul;23(7):567-576 [FREE Full text] [doi: 10.1089/tmj.2016.0200] [Medline: 28067586]
36. Alami H, Gagnon MP, Wootton R, Fortin JP, Zanaboni P. Exploring factors associated with the uneven utilization of telemedicine in Norway: a mixed methods study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):180 [FREE Full text] [doi: 10.1186/s12911-017-0576-4] [Medline: 29282048]
37. Bauer MS, Krawczyk L, Tuozzo K, Frigand C, Holmes S, Miller CJ, et al. Implementing and Sustaining Team-Based Telecare for Bipolar Disorder: Lessons Learned from a Model-Guided, Mixed Methods Analysis. Telemed J E Health 2018 Jan;24(1):45-53. [doi: 10.1089/tmj.2017.0017] [Medline: 28665773]
38. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-World Implementation of Video Outpatient Consultations at Macro, Meso, and Micro Levels: Mixed-Method Study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
39. Wade VA, Taylor AD, Kidd MR, Carati C. Transitioning a home telehealth project into a sustainable, large-scale service: a qualitative study. BMC Health Serv Res 2016 May 16;16(1):183 [FREE Full text] [doi: 10.1186/s12913-016-1436-0] [Medline: 27185041]
40. Lindsay JA, Kauth MR, Hudson S, Martin LA, Ramsey DJ, Daily L, et al. Implementation of video telehealth to improve access to evidence-based psychotherapy for posttraumatic stress disorder. Telemed J E Health 2015 Jun;21(6):467-472 [FREE Full text] [doi: 10.1089/tmj.2014.0114] [Medline: 25714664]
41. Latifi R, Dasho E, Merrell RC, Lopes M, Azevedo V, Bekteshi F, et al. Cabo Verde telemedicine program: initial results of nationwide implementation. Telemed J E Health 2014 Nov;20(11):1027-1034. [doi: 10.1089/tmj.2014.0026] [Medline: 25083737]
42. Lindsay JA, Day SC, Amspoker AB, Fletcher TL, Hogan J, Day G, et al. Personalized Implementation of Video Telehealth. Psychiatr Clin North Am 2019 Dec;42(4):563-574. [doi: 10.1016/j.psc.2019.08.001] [Medline: 31672207]
43. Darkins A. The growth of telehealth services in the Veterans Health Administration between 1994 and 2014: a study in the diffusion of innovation. Telemed J E Health 2014 Sep;20(9):761-768. [doi: 10.1089/tmj.2014.0143] [Medline: 25184945]
44. Rodgers EM. Diffusion of Innovations. 3rd edition. New York: Free Press; 1983.
45. Latifi R, Merrell RC, Doarn CR, Hadeed GJ, Bekteshi F, Lecaj I, et al. "Initiate-build-operate-transfer"--a strategy for establishing sustainable telemedicine programs in developing countries: initial lessons from the Balkans. Telemed J E Health 2009 Dec;15(10):956-969. [doi: 10.1089/tmj.2009.0084] [Medline: 19832055]
46. Telehealth Programs. Rockville, MD, USA; 2021 Jan. URL: https://www.hrsa.gov/rural-health/telehealth [accessed 2021-01-19]
47. Rycroft-Malone J. The PARIHS framework--a framework for guiding the implementation of evidence-based practice. J Nurs Care Qual 2004;19(4):297-304. [doi: 10.1097/00001786-200410000-00002] [Medline: 15535533]
48. Wootton R. Telemedicine in the National Health Service. J R Soc Med 1998 Dec 30;91(12):614-621 [FREE Full text] [doi: 10.1177/014107689809101202] [Medline: 10730107]
49. Crotty M, Killington M, van den Berg M, Morris C, Taylor A, Carati C. Telerehabilitation for older people using off-the-shelf applications: acceptability and feasibility. J Telemed Telecare 2014 Oct;20(7):370-376. [doi: 10.1177/1357633X14552382] [Medline: 25399997]
50. Birks M, Mills J. Grounded Theory: A Practical Guide. Los Angeles, CA, USA: Sage Publications; 2011.
51. Waterman H, Marshall M, Noble J, Davies H, Walshe K, Sheaff R, et al. The role of action research in the investigation and diffusion of innovations in health care: the PRIDE project. Qual Health Res 2007 Mar;17(3):373-381. [doi: 10.1177/1049732306298976] [Medline: 17301345]
52. Zanaboni P, Knarvik U, Wootton R. Adoption of routine telemedicine in Norway: the current picture. Glob Health Action 2014;7:22801 [FREE Full text] [doi: 10.3402/gha.v7.22801] [Medline: 24433942]
53. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11:33 [FREE Full text] [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]

54. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
55. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
56. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
57. Ritchie MJ, Dollar KM, Miller CJ, Smith JL, Oliver KA, Kim B, et al. Using Implementation Facilitation to Improve Care. Implementation Facilitation Training Manual (Version 3). Washington, DC, USA: Veterans Health Administration; 2020. URL: https://www.queri.research.va.gov/tools/implementation/Facilitation-Manual.pdf [accessed 2021-01-19]
58. Øvretveit J, Garofalo L, Mittman B. Scaling up improvements more quickly and effectively. Int J Qual Health Care 2017 Dec 01;29(8):1014-1019. [doi: 10.1093/intqhc/mzx147] [Medline: 29177491]
59. Barker PM, Reid A, Schall MW. A framework for scaling up health interventions: lessons from large-scale improvement initiatives in Africa. Implement Sci 2016 Jan 29;11:12 [FREE Full text] [doi: 10.1186/s13012-016-0374-x] [Medline: 26821910]
60. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 02;8(1):117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]
61. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018 May 14;16(1):66 [FREE Full text] [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
62. Nemeth C, Wears R, Woods D, Hollnagel E, Cook R. Minding the Gaps: Creating Resilience in Health Care. In: Henriksen K, editor. Advances in Patient Safety: New Directions and Alternative Approaches (Vol 3: Performance and Tools). Rockville, MD, USA: Agency for Healthcare Research and Quality; 2008:268-278.
63. Wade V, Eliott J. The role of the champion in telehealth service development: a qualitative analysis. J Telemed Telecare 2012 Dec;18(8):490-492. [doi: 10.1258/jtt.2012.gth115] [Medline: 23209264]
64. Tuot DS, Leeds K, Murphy EJ, Sarkar U, Lyles CR, Mekonnen T, et al. Facilitators and barriers to implementing electronic referral and/or consultation systems: a qualitative study of 16 health organizations. BMC Health Serv Res 2015 Dec 19;15:568 [FREE Full text] [doi: 10.1186/s12913-015-1233-1] [Medline: 26687507]
65. Bradford N, Caffery L, Smith A. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. Rural Remote Health 2016;16(4):4268 [FREE Full text] [Medline: 27817199]
66. Joseph V, West RM, Shickle D, Keen J, Clamp S. Key challenges in the development and implementation of telehealth projects. J Telemed Telecare 2011 Nov 19;17(2):71-77. [doi: 10.1258/jtt.2010.100315] [Medline: 21097563]
67. Osman MA, Schick-Makaroff K, Thompson S, Bialy L, Featherstone R, Kurzawa J, et al. Barriers and facilitators for implementation of electronic consultations (eConsult) to enhance access to specialist care: a scoping review. BMJ Glob Health 2019;4(5):e001629 [FREE Full text] [doi: 10.1136/bmjgh-2019-001629] [Medline: 31565409]
68. Moroz I, Archibald D, Breton M, Cote-Boileau E, Crowe L, Horsley T, et al. Key factors for national spread and scale-up of an eConsult innovation. Health Res Policy Syst 2020 Jun 03;18(1):57 [FREE Full text] [doi: 10.1186/s12961-020-00574-0] [Medline: 32493357]
69. Shaw J, Shaw S, Wherton J, Hughes G, Greenhalgh T. Studying Scale-Up and Spread as Social Practice: Theoretical Introduction and Empirical Case Study. J Med Internet Res 2017 Jul 07;19(7):e244 [FREE Full text] [doi: 10.2196/jmir.7482] [Medline: 28687532]
70. Leykum LK, Pugh J, Lawrence V, Parchman M, Noël PH, Cornell J, et al. Organizational interventions employing principles of complexity science have improved outcomes for patients with Type II diabetes. Implement Sci 2007 Aug 28;2(1):28 [FREE Full text] [doi: 10.1186/1748-5908-2-28] [Medline: 17725834]
71. Easterbrook P, Gopalan R, Berlin J, Matthews D. Publication bias in clinical research. The Lancet 1991 Apr;337(8746):867-872. [doi: 10.1016/0140-6736(91)90201-y]
72. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 03;320(7248):1517-1520 [FREE Full text] [doi: 10.1136/bmj.320.7248.1517] [Medline: 10834899]
73. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
74. Lin M, Yuan W, Huang T, Zhang H, Mai J, Wang J. Clinical effectiveness of telemedicine for chronic heart failure: a systematic review and meta-analysis. J Investig Med 2017 Jun;65(5):899-911. [doi: 10.1136/jim-2016-000199] [Medline: 28330835]
75. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]

76. Mounessa JS, Chapman S, Braunberger T, Qin R, Lipoff JB, Dellavalle RP, et al. A systematic review of satisfaction with teledermatology. J Telemed Telecare 2018 May;24(4):263-270. [doi: 10.1177/1357633X17696587] [Medline: 28350281]
77. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
78. Tzelepis F, Paul CL, Williams CM, Gilligan C, Regan T, Daly J, et al. Real-time video counselling for smoking cessation. Cochrane Database Syst Rev 2019 Oct 29;2019(10):1 [FREE Full text] [doi: 10.1002/14651858.CD012659.pub2] [Medline: 31684699]

Abbreviations

CVT: clinical video technology
CFIR: Consolidated Framework for Implementation Research
IBOT: initiate-build-operate-transfer
i-PARIHS: integrated Promoting Action on Research Implementation in Health Services
IT: information technology
MeSH: Medical Subject Headings
MH: Major Headings
NASSS: Nonadoption, Abandonment, and challenges to the Scale-up, Spread, and Sustainability
PARIHS: Promoting Action on Research Implementation in Health Services
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
PROSPERO: International Prospective Register of Systematic Reviews
PTSD: posttraumatic stress disorder
RE-AIM: Reach, Efficacy, Adoption, Implementation, and Maintenance
VA: Veterans Affairs
VHA: Veterans Health Administration
VoIP: Voice over Internet Protocol
VTH: video-to-home

Edited by R Kukafka; submitted 24.08.20; peer-reviewed by J Ledger, R Symonds, F Tzelepis, D Tuot; comments to author 25.09.20; revised version received 27.11.20; accepted 17.12.20; published 26.01.21.

Please cite as:
James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE
Spread, Scale-up, and Sustainability of Video Consulting in Health Care: Systematic Review and Synthesis Guided by the NASSS Framework
J Med Internet Res 2021;23(1):e23775
URL: http://www.jmir.org/2021/1/e23775/
doi:10.2196/23775
PMID:33434141

Original Paper

Adherence of Mobile App-Based Surveys and Comparison With Traditional Surveys: eCohort Study

Chathurangi H Pathiravasan[1], MS, PhD; Yuankai Zhang[1], BS; Ludovic Trinquart[1], MPH, PhD; Emelia J Benjamin[2,3], MD, ScM; Belinda Borrelli[4], PhD; David D McManus[5,6], MD, ScM; Vik Kheterpal[7], MD; Honghuang Lin[8], PhD; Mayank Sardana[9], MBBS; Michael M Hammond[3], MPH, MD; Nicole L Spartano[10], PhD; Amy L Dunn[3], BS; Eric Schramm[7], PharmD; Christopher Nowak[7], PhD; Emily S Manders[3], BS; Hongshan Liu[3], BS; Jelena Kornej[3], MD, MSc; Chunyu Liu[1], PhD; Joanne M Murabito[11], MD, ScM

[1]Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States

[2]Section of Preventive Medicine and Epidemiology and Cardiovascular Medicine, Department of Medicine, and Department of Epidemiology, Boston University Schools of Medicine and Public Health, Boston, MA, United States

[3]Boston University's and National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, United States

[4]Center for Behavioral Science Research, Department of Health Policy & Health Services Research, Boston University Henry M Goldman School of Dental Medicine, Boston, MA, United States

[5]Cardiology Division, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[6]Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[7]Care Evolution, Ann Arbor, MI, United States

[8]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, United States

[9]Cardiology Division, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[10]Section of Endocrinology, Diabetes, Nutrition, and Weight Management, Boston University School of Medicine, Boston, MA, United States

[11]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, United States

Corresponding Author:
Joanne M Murabito, MD, ScM
Section of General Internal Medicine
Department of Medicine
Boston University School of Medicine
Crosstown 2
201 Massachusetts Ave
Boston, MA, 02118
United States
Phone: 1 508 935 3461
Email: murabito@bu.edu

Abstract

Background: eCohort studies offer an efficient approach for data collection. However, eCohort studies are challenged by volunteer bias and low adherence. We designed an eCohort embedded in the Framingham Heart Study (eFHS) to address these challenges and to compare the digital data to traditional data collection.

Objective: The aim of this study was to evaluate adherence of the eFHS app-based surveys deployed at baseline (time of enrollment in the eCohort) and every 3 months up to 1 year, and to compare baseline digital surveys with surveys collected at the research center.

Methods: We defined adherence rates as the proportion of participants who completed at least one survey at a given 3-month period and computed adherence rates for each 3-month period. To evaluate agreement, we compared several baseline measures obtained in the eFHS app survey to those obtained at the in-person research center exam using the concordance correlation coefficient (CCC).

Results: Among the 1948 eFHS participants (mean age 53, SD 9 years; 57% women), we found high adherence to baseline surveys (89%) and a decrease in adherence over time (58% at 3 months, 52% at 6 months, 41% at 9 months, and 40% at 12 months). eFHS participants who returned surveys were more likely to be women (adjusted odds ratio [aOR] 1.58, 95% CI

1.18-2.11) and less likely to be smokers (aOR 0.53, 95% CI 0.32-0.90). Compared to in-person exam data, we observed moderate agreement for baseline app-based surveys of the Physical Activity Index (mean difference 2.27, CCC=0.56), and high agreement for average drinks per week (mean difference 0.54, CCC=0.82) and depressive symptoms scores (mean difference 0.03, CCC=0.77).

Conclusions: We observed that eFHS participants had a high survey return at baseline and each 3-month survey period over the 12 months of follow up. We observed moderate to high agreement between digital and research center measures for several types of surveys, including physical activity, depressive symptoms, and alcohol use. Thus, this digital data collection mechanism is a promising tool to collect data related to cardiovascular disease and its risk factors.

(J Med Internet Res 2021;23(1):e24773) doi:10.2196/24773

KEYWORDS

eCohort; mobile health; mHealth; smartphone; survey; app; Framingham Heart Study; adherence; agreement; cardiovascular disease

Introduction

eCohorts use new sensor devices and smartphone technology for longitudinal research data collection [1]. Technology permits the identification of digital biomarkers of health and improvement in health-related behaviors [2-5]. Mobile apps may be a promising and feasible tool for health interventions [6,7], and most previous studies have shown that mobile health (mHealth) plays an important role in promoting behavior change for children, adolescents, and young adults [8,9]. However, eCohorts are challenged by low adherence [10,11] and may yield substantial volunteer bias [12], raising concerns about the generalizability of study findings. For example, in the MyHeart Counts Cardiovascular Health Study [11] that examined the feasibility of a smartphone-based assessment of physical activity, less than 10% of enrolled participants completed the full 7 days of physical activity data. Because of low adherence rates and sampling bias, it remains unclear how these previous study results will reflect the whole community.

Establishing the validity and reliability of new electronic data collection methods is required before deploying digital technology in epidemiology settings [13]. A few studies have investigated the equivalence of questionnaires administered on different electronic devices versus traditional paper data collection [14-17]. The electronic modes in these studies included a tablet, touchscreen, interactive voice response system, and personal digital assistant. A few studies attempted to create smartphone apps for specific clinical use, which demonstrated scientific validity [18,19]. However, reliability assessment of mobile app surveys remains scarce, especially for cardiovascular phenotypes.

To integrate digital and mHealth data into a traditional longitudinal cohort, we leveraged an in-person examination as part of the Framingham Heart Study (FHS) to enroll participants into an eCohort (eFHS) using a new smartphone app, digital blood pressure cuff, and smartwatch [20]. Embedding our eCohort in the FHS allowed us to compare the digital measures obtained from a smartphone with the same surveys obtained during in-person clinical examinations at the FHS Research Center using research protocols [21]. To that end, the objective of this study was threefold: (1) to determine the app-based survey adherence over a 12-month period, (2) to compare baseline eFHS app survey measurements to research center measures, and (3) to examine the association of periodic app-based survey measures across different time points. We hypothesized that embedding eFHS in the FHS and leveraging the in-person exam to enroll participants would result in high app-based survey adherence at baseline that would decrease over the 1-year follow-up period. We also hypothesized that app-based surveys would be comparable with surveys collected at the research center.

Methods

Study Sample

Participants of the FHS Third Generation (Gen 3) cohort (n=4095), Omni Group 2 (n=410), and New Offspring Spouse (n=103) were recruited from 2002 to 2005, and underwent periodic research examinations every 6 to 8 years [21]. We leveraged exam 3 (2016 to 2019) to invite English-speaking FHS participants who owned a smartphone (including iPhone 4S or higher with at least iOS 8.2, or an Android phone as of October 30, 2017) to enroll in the eFHS. The eFHS study was approved by the Institutional Review Board at Boston University Medical Center. Beginning in June 2016, at the time of the in-person research examination, participants were invited to download the eFHS smartphone app from the Apple App Store. The participants were not required to register and complete the app-based surveys at baseline at the research center. Some participants chose to register after leaving the research center. Participation in the eFHS was voluntary and participants were not incentivized for participating. All participants were provided with a written protocol that includes information of how to download the app, enter registration information, sign the electronic consent form, and enable notifications on the phone (Multimedia Appendix 1). Participants reached the first screen upon logging into the app after registration. The list of surveys was organized by due date and displayed on the survey screen (Multimedia Appendix 1). Participants received different types of notifications such as welcome messages on enrollment to the study, notifications when new surveys became available, reminder notifications to complete surveys, and thank you messages after completing all surveys (Multimedia Appendix 2). Collected survey data in this study were pushed to a secure cloud server and transmitted to FHS Research Center servers. Among the 3521 participants (Multimedia Appendix 3) who came to the research center (from April 2016 to April 2019), we excluded 1370 participants who were ineligible (did not own a smartphone), did not consent (owned a smartphone but declined participation), or had incompatible phones, and those who enrolled in eFHS but had less than a 12-month follow-up period (n=203).

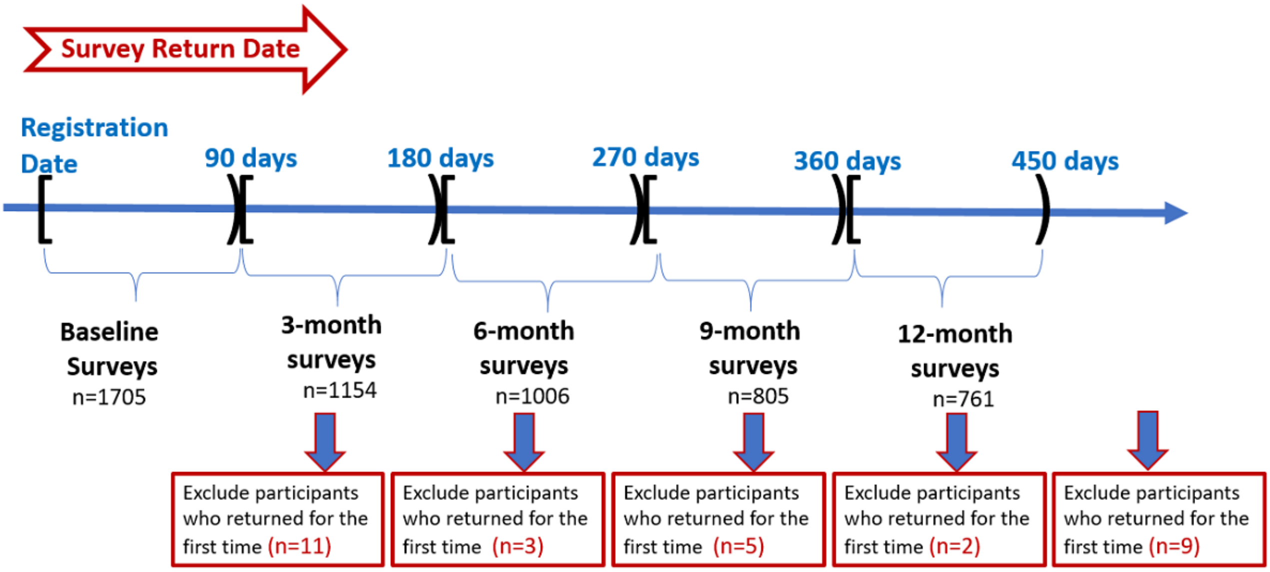
eFHS Smartphone App

Health surveys were distributed at baseline (app download) and at 3-month intervals for 1 year [20] to collect sociodemographic data and variables related to cardiovascular disease (CVD). Corresponding surveys were administered in the research center by trained research technicians or physician/nurse practitioners, except for the sociodemographic survey and health survey questions, which were self-administered. A total of 22 surveys were deployed on the smartphone over the 12-month period after enrollment. Nine baseline surveys for self-reported data were deployed in the following order: sociodemographic information, smoking, medications and self-reported risk factors, baseline CVD history, baseline non-CVD medical history, physical activity, alcohol consumption, health survey, and depressive symptoms (assessed with the Center for Epidemiologic-Studies Depression Scale [CES-D]) [22]. Among the nine baseline surveys, the physical activity survey was deployed at each 3-month interval after registration. A medical history update, depressive symptoms (CES-D), and health survey were gathered at 6 months and 12 months. Surveys for medication use, self-reported risk factors, and smoking and alcohol consumption were collected at 12 months after the baseline survey. A short description of each survey (Multimedia Appendix 4) and the number of steps to complete each survey were provided to the participants. The first step provided the purpose of the survey, intermediate steps contained one or more survey questions, and the last step thanked participants for completing the survey (see Multimedia Appendix 1 for screenshots of some steps of the Physical Activity Index [PAI] survey). Detailed descriptions of screens or steps of each of 10 different eFHS app surveys can be found in our previous report [20].

eFHS Smartphone Survey 3-Month Interval

We evaluated adherence by assessing the return of baseline survey data within 90 days of registration. To distinguish returned survey data from baseline or additional follow-up surveys, we used a time window approach (Figure 1). Thirty participants were excluded because they returned surveys for the first time after 90 days from the registration. We considered the survey data as baseline surveys if participants returned data between 0 to 89 days from registration. If the surveys were returned between 90 to 179 days from the registration date, we considered the data as 3-month surveys. Similarly, we considered surveys as 6-month, 9-month, and 12-month survey data if participants returned their surveys within 180-269 days, 270-359 days, and 360-449 days from the registration dates, respectively.

Figure 1. The time window approach for separating baseline and other follow-up surveys (here, completing a survey means completing 75% of questions in the given survey).



Survey Return Time and Touch Time

To evaluate survey completion time, we examined the survey return time and touch time. The survey return time was calculated by considering the time between the deployment of the survey and the return of a survey. The touch time was calculated by taking the time between the start and return of a survey. We computed the median and IQR of the touch time and survey return time for each type of survey (Multimedia Appendix 5). Furthermore, we calculated the step time (time taken to complete each step) and question time (time taken to complete a question).

Survey Adherence

To minimize frustration, we allowed eFHS participants to skip questions in a given survey. We defined survey completion (completing one survey) if a participant completed 75% of all questions in a given survey (Multimedia Appendix 6). We used two methods for calculating survey adherence. We first calculated the proportion of individuals who completed at least one survey at a given 3-month time window. The second method calculated the proportion of individuals who completed all surveys at a given 3-month time window (Multimedia Appendix 7).

Statistical Analysis

We compared the characteristics of FHS participants who provided consent for eFHS with those of participants who declined or were not eligible for this study. All characteristic variables were collected at research center health examination 3. We used Student *t* tests for continuous variables and χ^2 tests for categorical variables. Among the participants who enrolled in eFHS, we compared the characteristics of participants who returned smartphone app surveys with those of participants who did not return surveys. We used a multivariable (adjusted) logistic regression model, which included an indicator variable (to denote the eFHS participants who returned surveys and those who did not return) as the dependent variable adjusting for age, sex, current smoking, and highest education level.

Surveys collected at the research center health examination and the eFHS app-based surveys provided two sets of measurements for sociodemographic and medical information. To evaluate the agreement of digital survey measures, we compared surveys at eFHS baseline to surveys collected from research center examination 3 using the concordance correlation coefficient (CCC) [23,24] and Bland-Altman plots [25,26]. We used the measurements in the research center as the gold standard for the Bland-Altman analysis. We investigated three different types of surveys deployed at different time intervals: PAI, depressive symptoms surveys (CES-D score), and alcohol consumption surveys. The PAI surveys were deployed every 3 months, CES-D surveys were deployed every 6 months, and the alcohol surveys were deployed at baseline and at 12 months. The surveys collect different health behaviors and mood information that might be reported differently when administered by a trained examiner vs self-reported on a smartphone app.

We calculated PAI as a weighted composite score [27-29] of activity levels with corresponding weights of 1 (for sleep), 1.1 (for sedentary), 1.5 (for slight), 2.4 (for moderate), and 5 (for heavy activities). The number of hours for each physical activity variable (sleep, sedentary, slight, moderate, and heavy) summed to 24 hours. If an individual had one variable missing, we imputed the missing value with 24 minus the sum of hours from the other four variables (n=77).

We analyzed depressive symptoms with two variables: a continuous variable that was the summation of the individual CES-D scores and a binary variable that was defined as 1 if the sum of the CES-D score was ≥16 and 0 otherwise [30,31]. To calculate the continuous CES-D variable, we considered only participants who answered all 20 questions. We found that 72 participants skipped at least one question. We imputed the 61 missing values using the following rule [32]: if more than 5 items were missing, all observations were considered as missing (n=11); if 1 to 5 items were missing, then the average value of the nonmissing items was multiplied by 20.

For alcohol use, we used average drinks per week calculated by the number of drinks per day times the number of days a participant had any type of alcoholic beverage per week.

We used the CCC for continuous measures and the Cohen κ coefficient [33,34] for categorical predictors for agreement analysis between measures from the same participants between eFHS app surveys and research center questionnaires.

We also examined whether three measures (PAI, the sum of CES-D, and average drinks per week) displayed any trend across time points. We used linear mixed models to compare means of PAI and depression symptom measures at each 3-month period. We used the paired *t* test to compare alcohol consumption at baseline and 12 months. We used the R program (version 3.6.1) for all statistical analyses and considered two-sided *P* values <.05 to indicate statistical significance.

Results

Study Sample and Survey Metrics

We included 1948 eFHS participants who met the eligibility criteria and had a follow-up time of 12 months or longer (Multimedia Appendix 3) for analysis. Compared to FHS participants who were not enrolled in eFHS, enrolled participants were more likely to be women, white, married, employed full time, completed bachelor or higher degrees, report excellent health, and have more favorable CVD risk factor levels (Table 1). Among the enrolled eFHS participants, 1735 participants returned surveys. Participants who returned surveys were more likely to be women (adjusted odds ratio [aOR] 1.58, 95% CI 1.18-2.11) and were less likely to be current smokers (Multimedia Appendix 8).

Table 1. Characteristics of participants in the eCohort Framingham Heart Study (eFHS) and Framingham Heart Study (FHS) participants not enrolled in eFHS.[a]

Variable	eFHS participants (n=1948)	FHS participants not enrolled in eFHS (n=1566)	*P* value
Age (years), mean (SD)	52.8 (8.7)	56.6 (9.8)	<.001
Female sex, n (%)	1109 (56.93)	782 (49.94)	<.001
Race, n (%)			.003
White	1813 (93.07)	1414 (90.29)	
Black	30 (1.54)	29 (1.85)	
Hispanic	45 (2.31)	60 (3.83)	
Asian	28 (1.44)	43 (2.75)	
Other	32 (1.64)	20 (1.28)	
BMI (kg/m^2), median (IQR)	27.3 (24.3-31.3)	28.2 (24.6-32.6)	<.001
Systolic blood pressure (mmHg), mean (SD)	119 (14)	121 (15)	<.001
Diastolic blood pressure (mmHg), mean (SD)	76 (8)	76 (9)	.33
Current smoking, n (%)[b]	108 (5.55)	125 (7.99)	.005
Diabetes mellitus, n (%)[c]	122 (6.29)	189 (12.36)	<.001
Hypertension, n (%)[d]	511 (26.26)	582 (37.26)	<.001
Atrial fibrillation, n (%)	36 (1.85)	52 (3.32)	.008
Prevalent cardiovascular disease, n (%)	67 (3.44)	97 (6.19)	<.001
Physical Activity Index score, median (IQR)	32.60 (30.10-35.50)	33 (30.20-36.40)	.004
Highest education level achieved, n (%)			<.001
Less than high school	13 (0.67)	35 (2.26)	
Completed high school	168 (8.66)	301 (19.46)	
Completed some college	467 (24.08)	441 (28.51)	
Bachelor's degree	741 (38.22)	479 (30.96)	
Graduate or professional degree	550 (28.37)	291 (18.81)	
Married, living as married, living with partner, n (%)[e]	1446 (74.73)	1002 (65.15)	<.001
Self-reported health as excellent, n (%)[f]	1414 (72.62)	913 (58.60)	<.001
Employed full time, n (%)[g]	1361 (70.19)	913 (59.32)	<.001

[a] Data reflect enrollment up to January 28, 2019.
[b] eFHS n=1947, FHS n=1564.
[c] eFHS n=1940, FHS n=1529.
[d] eFHS n=1946, FHS n=1562.
[e] eFHS n=1935, FHS n=1538.
[f] eFHS n=1947, FHS n=1558.
[g] eFHS n=1939, FHS n=1539.

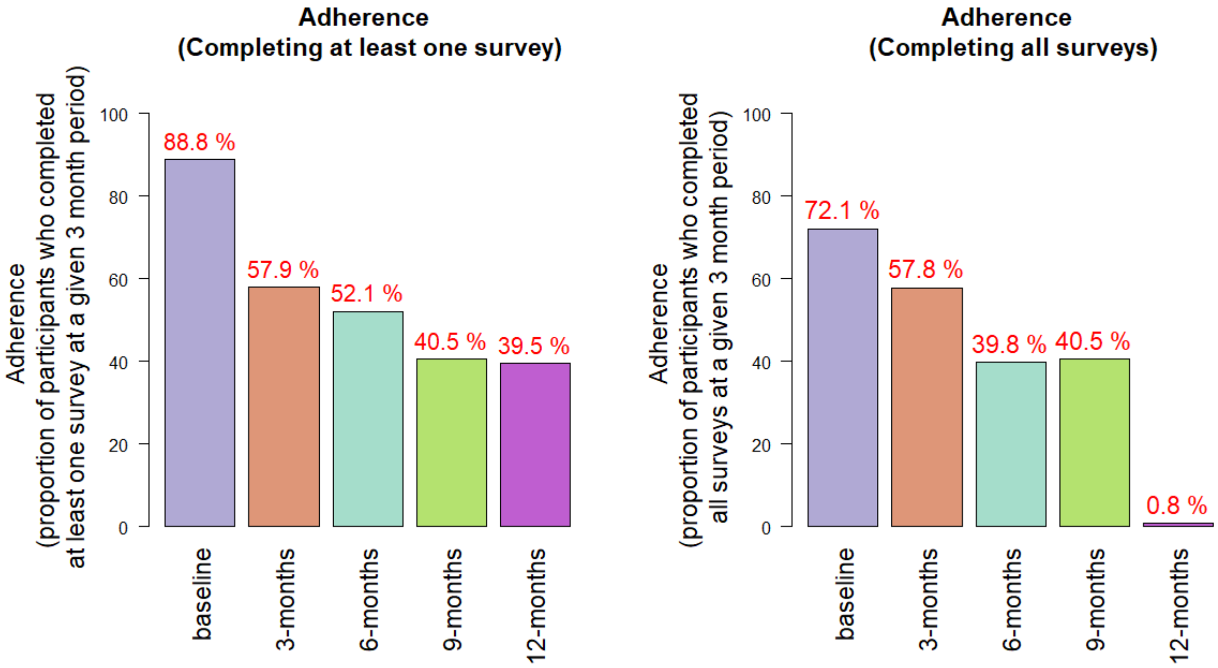
There were 1705 participants who returned baseline surveys within the 3-month interval (Multimedia Appendix 3). Survey return time, touch time, step time, and time spent for each question for each survey at each survey wave are presented in Multimedia Appendix 5. Participants returned baseline and follow-up surveys within 2 weeks with the exception of the 12-month medical history update that was returned within 30 days (Multimedia Appendix 9). The actual time taken to complete surveys (touch time) was less than 5 minutes.

The number of participants who returned a specific type of survey at each survey wave is displayed in Multimedia Appendix 10. Most participants completed 75% of the questions of at least one baseline survey (1704/1918, 88.84%; see Multimedia Appendix 7). Participants continued to return some surveys at 12 months (757/1918, 39.47%; see Multimedia Appendix 7) but only a few participants (28/1918, 1.46%; see Multimedia Appendix 10) completed the 12-month medical history update questionnaire.

Survey Adherence/Completion

Among participants who returned surveys, more than 85% completed more than 75% of the questions across all surveys at all time windows (Multimedia Appendix 6). Considering the proportion of individuals who completed at least one survey at a given 3-month window (Multimedia Appendix 7), eFHS participants had the highest adherence at baseline (89%) and the adherence decreased over time (58% at 3 months, 52% at 6 months, 41% at 9 months, and 40% at 12 months) (Figure 2). We observed similar adherence rates based on the proportion of participants who completed all surveys at a given 3-month period (72% at baseline, 58% at 3 months, 40% at 6 months, 41% at 9 months). At 12 months, adherence was reduced to 1% because most participants did not return the 12-month medical history update survey.

Figure 2. Proportion of eCohort Framingham Heart Study (eFHS) participants who completed at least one survey (left panel) and all surveys (right panel).

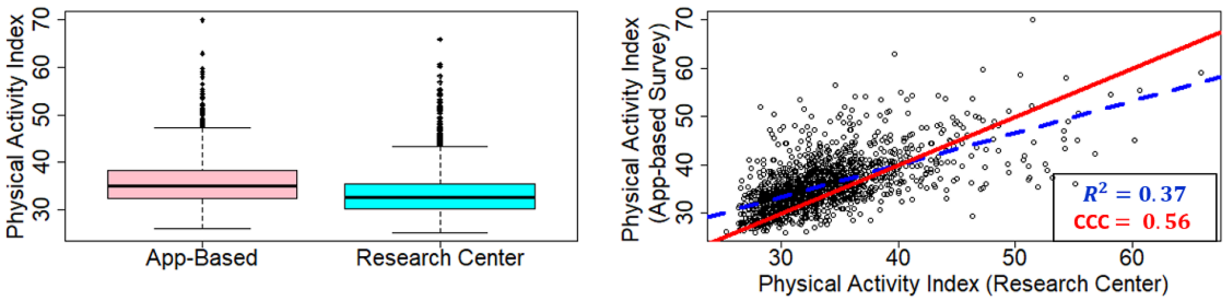


Comparison of Baseline eFHS App Survey With Research Center Measurements

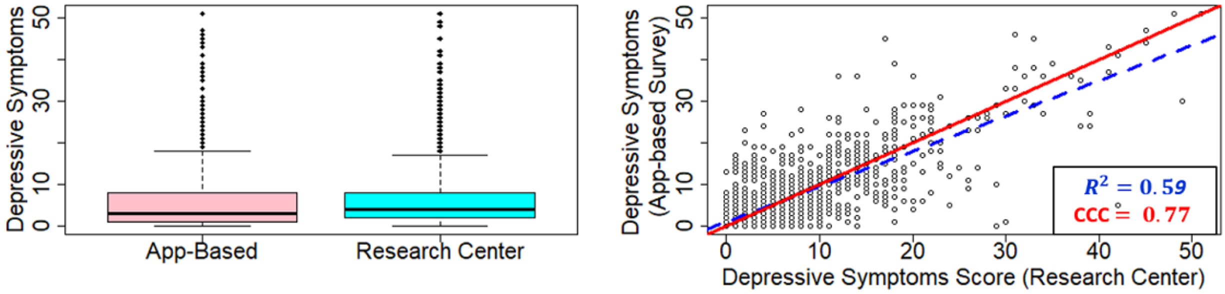
The baseline app-based survey data for PAI (n=1545), CES-D score (n=1628), and average alcohol consumption per week (n=1513) were used to compute the mean difference between the mobile app surveys and the questionnaires collected in the research center. The mobile app survey had a higher PAI compared to the respective in-research center exam (mean difference 2.27). We observed high agreement (CCC=0.82, 95% CI 0.81-0.84) between the two alcohol consumption measures and the two CES-D scores (CCC=0.77, 95% CI 0.75-0.79) (Figure 3). Moderate agreement was observed between the two PAI measures (CCC=0.56, 95% CI 0.52-0.59) and the binary depression variables (Cohen κ=0.51, 95% CI 0.44-0.58). The Bland-Altman plots of PAI, CES-D scores, and alcohol consumption showed that the spread of the difference increased with increasing mean of the observations (bias increased and variability was not consistent across the graph), likely reflecting that the distributions were skewed (Multimedia Appendix 11).

Figure 3. Comparison of baseline surveys: Physical Activity Index (PAI), depression symptoms scale (Sum of Center For Epidemiologic-Studies Depression Scale [CES-D] scores), and alcohol consumption (average drinks per week).

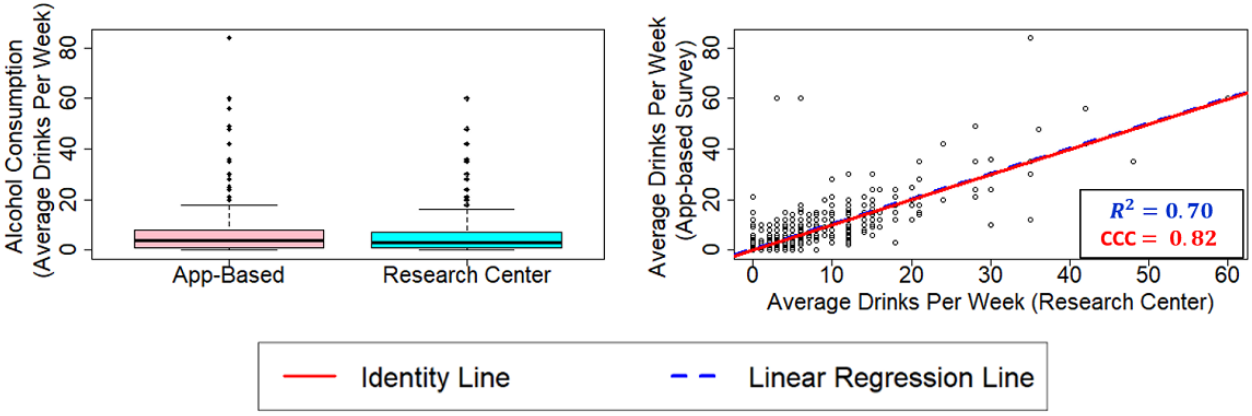
(a) Comparison Of Physical Activity Index



(b) Comparison Of Depressive Symptoms



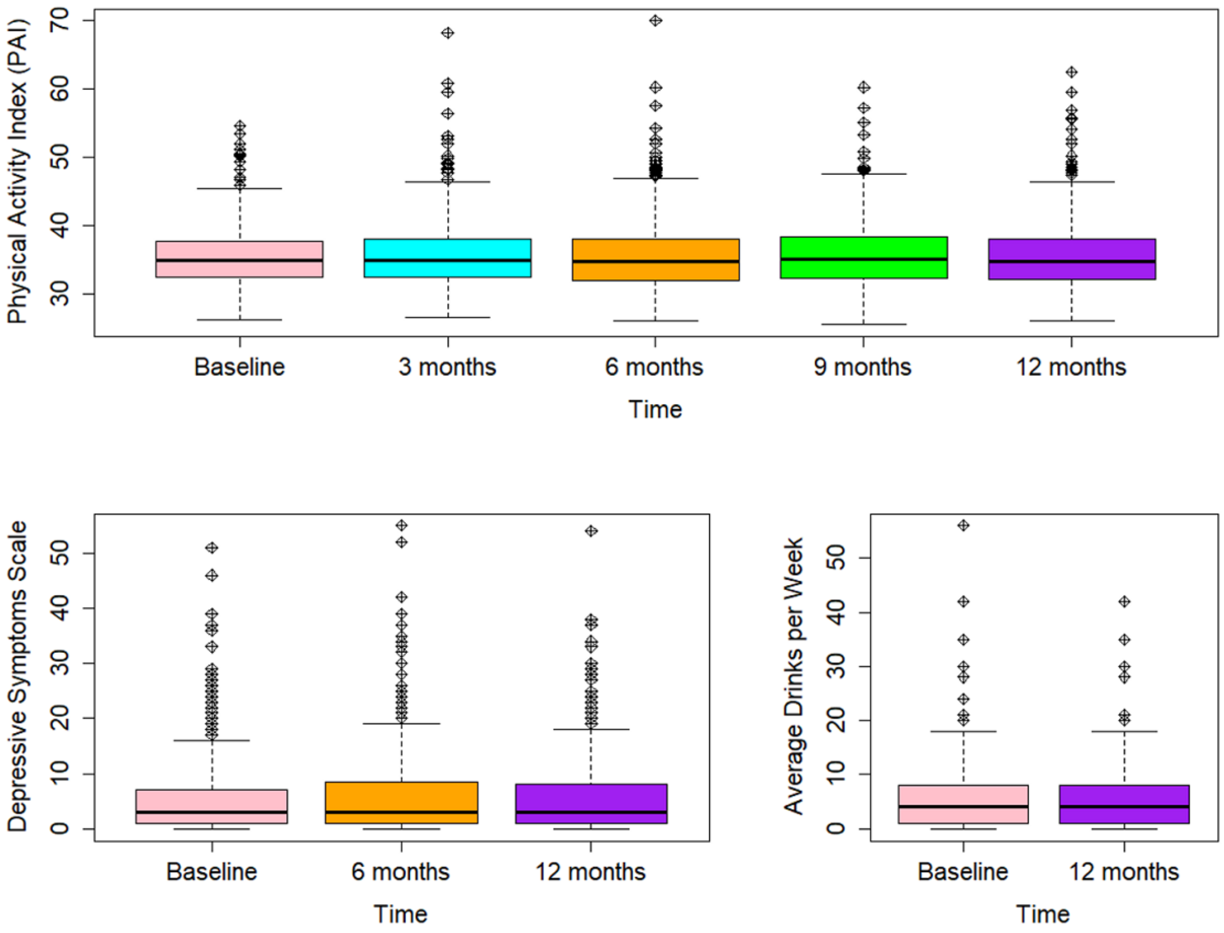
(c) Comparison Of Alcohol Consumption



Association of Periodic App-Based Survey Measures Across Different Time Points

Overall, 539 participants returned physical activity questionnaires at all five time points (baseline, 3 months, 6 months, 9 months, and 12 months), 644 participants returned depressive symptom surveys at all three time points (baseline, 6 months, and 12 months), and 613 participants returned all alcohol consumption surveys (baseline and 12 months). PAI (for all time effects compared to baseline) and alcohol consumption (mean difference of baseline to 12-month questionnaire –0.03, *P*=.82) were at similar levels across several time points, whereas the CES-D score slightly increased from baseline to 6 months (slope=1.01, *P*<.001) and from baseline to 12 months (slope=0.84, *P*<.001) (Figure 4).

Figure 4. Boxplots for Physical Activity Index (PAI), depressive symptoms scale (Center For Epidemiologic-Studies Depression Scale [CES-D] score), and average drinks per week across different time points.



Discussion

Principal Results

Our findings in this middle-aged community-based sample are threefold. First, eFHS study participants had high survey adherence at baseline. The majority of participants who returned surveys completed more than 75% of the survey questions across each 3-month period. Adherence decreased over the 12-month follow-up period but remained high compared with that reported in previous mHealth studies. Second, eFHS is embedded in the ongoing prospective FHS, providing us the opportunity to compare surveys collected on the mobile app to questionnaires collected in the research center using standardized protocols. Our data suggest that app-based surveys and research center questionnaires for physical activity, mood, and alcohol intake had moderate to high agreement. Finally, among the subgroup of participants who returned all follow-up surveys, reports of physical activity (PAI) and alcohol consumption were consistent across all 3-month time intervals over the 12-month follow up.

Comparison With Previous Studies

Although digital and mHealth technologies have great potential to improve health, it is essential to understand engagement challenges with such technologies, as most mHealth studies have observed substantial participant drop offs within a short time period [35]. For example, in the recent Asthma Mobile Health Study [10], more than 40,000 participants downloaded the app, whereas only 7500 participants enrolled in the study and only 175 (2.3%) participants continued to contribute data at the 6-month follow up. Similarly, a study of posttraumatic stress disorder initially had 166,800 participants download the app but only 26,110 (15.7%) users remained after 1 week [36]. The Health eHeart Study is another eCohort that combined the use of social media, smartphones, and wearables to study heart disease. In this groundbreaking cardiovascular study, 86% of consented participants completed at least one survey but 42% of the surveys had missing values [12]. eFHS participants demonstrated considerably higher engagement with a substantial number of participants returning completed surveys at the 12-month follow up (Multimedia Appendix 10). Several reasons may explain the high levels of engagement among eFHS participants. First, the eFHS is embedded in the FHS that consists of loyal participants; the original cohort was followed for more than 70 years, and the present cohort consists of their grandchildren who have been followed in three cycles since the early 2000s. Therefore, the eFHS participants have a long and positive relationship with the study and staff. The smartphone and associated study devices may further connect the participant to the research staff and allow participants to gain insight into their health that could potentially impact engagement [37].

Second, participants received positive notifications, including notifications of thanks and acknowledgement ("Thank you for completing your surveys. Your contribution is a vital part in our ongoing research efforts!"), which may have encouraged participation.

Recent studies have tested the equivalence between electronic and paper administration of patient-reported outcome measures [14,15]. These studies focused on acceptability of the different electronic data collection modes, including personal digital assistants using a tablet or personal computer and interactive voice response systems (automated telephone questionnaire), but did not consider mobile apps. Few studies tested the reliability and validity of mobile apps that were specifically designed for special research purposes. The Burden of Obstructive Lung Disease (BOLD) study was conducted to measure the burden of chronic obstructive lung disease; researchers compared the smartphone and paper-based data collection systems in rural Sudan [19]. A new smartphone app of the International Prostate Symptom Score was also developed and tested [18]. These studies demonstrated that smartphone technology worked well compared to paper-based data collection. Nevertheless, validation studies are limited, and more investigation is needed to recommend a mobile app as an effective method of data collection compared to traditional paper-based surveys. Our study demonstrated good agreement between app-based surveys and surveys collected in the research center using standard protocols across a range of measurements, including physical activity, mood, and alcohol intake. Participants reported a slightly higher PAI in the app-based surveys compared to the same questionnaires filled out at the research center. Another study demonstrated equivalence between the paper and smartphone versions of two scales of depression [38], but some electronic items (such as sad mood and trouble concentrating) were lower compared to items on the paper versions even though they had moderate to high consistency. A randomized study was conducted to compare paper versus electronic mode of delivery of a health and social behavior questionnaire [39]. The majority of mode differences were nonsignificant, but participants reported more exercise in the paper survey compared to the electronic survey. This contrasts with our findings in this study as we observed higher PAI scores in app-based surveys. However, smartphone app surveys can be an efficient method to collect cardiovascular risk factor data.

Participants volunteered in the study and there was no mandatory requirement of answering questions in the smartphone app. Adherence varied across different types of surveys. The difficulty of the survey may impact long-term adherence. For example, we observed adherence rates of 40% even after 12 months of follow up for most surveys (n=757), but lower completion rates were obtained for the 12-month medical history update survey (n=28). It is unclear why participants did not complete the medical history update survey. However, this survey was longer than the other surveys in the eFHS app, and compared to other surveys, the medical history update survey contained many open-ended questions that may have required more time and effort to complete. Indeed, unrestricted or open-ended questions have been reported as the major challenge of app-based surveys [19,40].

We conducted a pilot study using two distinct enrollment methods to compare adherence for device use and two internet surveys. The two methods were on-site support (n=101) and remote (n=93) enrollment [41]. The baseline core internet survey consisted of 34 separate parts for self-reported health outcomes, which was completed at home after enrollment and consent. To address the overall study performance, an end-of-study survey was emailed to the participants at study termination. On-site support increased the participation and the initial rate of device use compared to remote support. The pilot study also demonstrated that on-site research center visit was associated with higher adherence to the end-of-study survey. However, on-site research center visit was associated with lower adherence to the internet survey at baseline compared to the remote arm. Thus, the pilot study suggests that in-person contact may not be as important for studies designed to deliver only surveys. We observed higher adherence to the baseline app-based survey in the eFHS than reported for other eCohorts. The eFHS is embedded in the FHS, which consists of loyal participants, potentially leading to higher baseline app-based adherence in the eFHS [37].

A new, enhanced version of the app was implemented in February 2019, which includes an interactive health dashboard to promote participant engagement and facilitate survey adherence. We did not uniformly collect feedback from users, but we are planning to assess usability of the new enhanced version of the app with the Mobile App Rating Scale (MARS) [42] as a separate study.

Insights and responses from participants may help to identify better strategies to improve adherence in mHealth studies [43]. Monitoring and feedback, reminders, goal setting, social support features, and rewards are tools that can be used to improve short- and long-term engagement [44-46]. In particular, the just-in-time adaptive intervention, which aims to provide a proper amount of support at the appropriate time when it is needed, may lead to a change in health behavior [47]. In future work, we plan to establish an advisory panel of eFHS participants to assist with co-design of the app and engagement methods [48].

Strengths and Limitations

The eFHS is a large community-based cohort designed to study CVD and other risk factors. It is embedded within the ongoing traditional FHS, and the study design provided the opportunity to compare FHS participants who consented vs participants who declined. It also provided the opportunity to compare baseline app surveys and surveys collected in the research center. Importantly, it provided the opportunity to compare eFHS participants to FHS participants who declined enrollment in eFHS to understand the generalizability of the sample.

Our study has several limitations. eFHS included predominantly white, educated, and healthy participants who own smartphones, most of whom resided in the New England region of the United States, thus limiting the ability to generalize the findings from this study to other racial and ethnic groups, individuals with less than high school education, smartphone-naïve, disease-based samples, and other regions/countries. Some participants downloaded the app (n=203) and enrolled in eFHS after January 31, 2019. Since these participants did not have the opportunity to participate in the study for at least 1 year at the time that we began the analysis, we excluded these participants from this study. There were slight differences in the characteristics (Multimedia Appendix 12) of eFHS enrollees who had a greater than 1-year follow-up time vs participants who had a less than 1-year follow-up time (women: 57% vs 50%, married: 75% vs 66%, and median BMI: 27 vs 29), which may not influence the generalizability of the current study. We allowed 90 days to complete baseline app-based surveys, but the participants returned digital surveys within a couple of weeks from the enrollment. The median survey return time varied from 1.35 to 2.40 days and the IQR was about 7 days (Multimedia Appendix 5). Therefore, the survey results relying on self-reported CVD risk factors may not differ according to the time of completing the survey, which would not affect the comparison results of research center measurements (time at enrollment). Moreover, since our study was observational, we cannot exclude residual confounding factors and cannot establish causal relations in our observations of variation in characteristics associated with adherence. A subset of eFHS participants (n=655) were enrolled in a randomized controlled messaging trial (ClinicalTrials.gov NCT03516019) [49] designed to test the impact of personalized notifications on device use. However, the notifications did not impact app-based survey adherence. Recent digital health intervention studies were found to be cost-effective in the management of CVD [50,51]. However, an economic evaluation of digital data was beyond the scope of this study.

Conclusions

We observed high adherence to baseline surveys with a substantial proportion of participants continuing to complete surveys at the 12-month follow up, indicating that the eFHS app may be a promising tool to collect data. App-based surveys were comparable to the research center–administered questionnaires. Therefore, the eFHS app may serve as a reliable data collection mechanism. Further exploration is needed to understand the reasons for the higher PAI in the app-based surveys compared to the surveys administered at the research center.

Acknowledgments

This study was supported by an award from the Robert Wood Johnson Foundation (number 74624) and a grant from the National Heart Lung and Blood Institute (R01HL141434). The FHS is supported by a contract from the National Heart, Lung, and Blood Institute (principal investigator VR; 75N92019D00031), and investigator time from the following grants: R01HL126911 (EB), 2R01 HL092577 (EB), 1R01AG066010 (EB), American Heart Association, 18SFRN34110082 (EB), 2U54HL120163 (EB), R01HL126911 (DM), R01HL137734 (DM), R01HL137794 (DM), R01HL13660 (DM), U54-HL 143541 (DM).

Conflicts of Interest

JM was a guest lecturer at Merck Research Laboratories. DM has received research support from Apple Computer, Bristol-Myers Squibb, Boehringer-Ingelheim, Pfizer, Flexcon, Samsung, Philips Healthcare, and Biotronik, and has received consultancy fees from Bristol-Myers Squibb, Pfizer, Flexcon, Boston Biomedical Associates, and Rose Consulting. DM also declares financial support for serving on the Steering Committee for the GUARD-AF study (NCT04126486) and Advisory Committee for the Fitbit Heart Study (NCT04176926). VK is principal, and CN and EM are employees of CareEvolution, Inc, a health care technology company. As of 2020, EB was an uncompensated member of the MyHeartLab Steering Committee, a principal investigator–initiated study from Samsung to University of California San Francisco (principal investigator, Jeffrey Olgin, MD). Other authors have no relevant disclosures.

Multimedia Appendix 1

(a) eFHS app loading screen and registration steps. (b) The first screen with the list of surveys and steps within the physical activity questionnaire.

[PDF File (Adobe PDF File), 384 KB - jmir_v23i1e24773_app1.pdf]

Multimedia Appendix 2

Notifications.

[PNG File , 48 KB - jmir_v23i1e24773_app2.png]

Multimedia Appendix 3

eFHS cohort development diagram for the study (app-based survey returns).

[PNG File , 34 KB - jmir_v23i1e24773_app3.png]

Multimedia Appendix 4

Survey description in the eFHS participant protocol.

[PDF File (Adobe PDF File), 278 KB - jmir_v23i1e24773_app4.pdf]

Multimedia Appendix 5
Survey summary.
[PDF File (Adobe PDF File), 30 KB - jmir_v23i1e24773_app5.pdf]

Multimedia Appendix 6
Determining the threshold for survey completion.
[PDF File (Adobe PDF File), 33 KB - jmir_v23i1e24773_app6.pdf]

Multimedia Appendix 7
Number and proportion of eligible eFHS participants completing at least one, all, or none of the periodic app-based surveys within the 3-month window.
[PDF File (Adobe PDF File), 185 KB - jmir_v23i1e24773_app7.pdf]

Multimedia Appendix 8
Characteristics of participants who returned surveys vs participants who did not return surveys.
[PDF File (Adobe PDF File), 293 KB - jmir_v23i1e24773_app8.pdf]

Multimedia Appendix 9
Survey return and touch time.
[PDF File (Adobe PDF File), 208 KB - jmir_v23i1e24773_app9.pdf]

Multimedia Appendix 10
Number of participants who completed 75% of questions of each survey over the 12-month period.
[PDF File (Adobe PDF File), 185 KB - jmir_v23i1e24773_app10.pdf]

Multimedia Appendix 11
Bland-Altman plots (mobile-based survey measurements vs research center measurements) of Physical Activity Index (PAI), depressive symptoms, and alcohol consumption.
[PNG File , 59 KB - jmir_v23i1e24773_app11.png]

Multimedia Appendix 12
Characteristics of eFHS enrollees by follow-up time.
[PDF File (Adobe PDF File), 288 KB - jmir_v23i1e24773_app12.pdf]

References

1. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011. [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
2. Huang Z, Tan E, Lum E, Sloot P, Boehm BO, Car J. A Smartphone App to Improve Medication Adherence in Patients With Type 2 Diabetes in Asia: Feasibility Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Sep 12;7(9):e14914 [FREE Full text] [doi: 10.2196/14914] [Medline: 31516127]
3. Park JYE, Li J, Howren A, Tsao NW, De Vera M. Mobile Phone Apps Targeting Medication Adherence: Quality Assessment and Content Analysis of User Reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
4. Hsieh W, Su YC, Han H, Huang M. A Novel mHealth Approach for a Patient-Centered Medication and Health Management System in Taiwan: Pilot Study. JMIR Mhealth Uhealth 2018 Jul 03;6(7):e154 [FREE Full text] [doi: 10.2196/mhealth.9987] [Medline: 29970356]
5. Morita PP, Yeung MS, Ferrone M, Taite AK, Madeley C, Stevens Lavigne A, et al. A Patient-Centered Mobile Health System That Supports Asthma Self-Management (breathe): Design, Development, and Utilization. JMIR Mhealth Uhealth 2019 Jan 28;7(1):e10956 [FREE Full text] [doi: 10.2196/10956] [Medline: 30688654]
6. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
7. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]

8. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: A systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
9. Radovic A, Badawy SM. Technology Use for Adolescent Health and Wellness. Pediatrics 2020 May;145(Suppl 2):S186-S194. [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
10. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
11. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of Obtaining Measures of Lifestyle From a Smartphone App: The MyHeart Counts Cardiovascular Health Study. JAMA Cardiol 2017 Jan 01;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
12. Guo X, Vittinghoff E, Olgin JE, Marcus GM, Pletcher MJ. Volunteer Participation in the Health eHeart Study: A Comparison with the US Population. Sci Rep 2017 May 16;7(1):1956. [doi: 10.1038/s41598-017-02232-y] [Medline: 28512303]
13. Roger VL, Boerwinkle E, Crapo JD, Douglas PS, Epstein JA, Granger CB, et al. Strategic transformation of population studies: recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. Am J Epidemiol 2015 Mar 15;181(6):363-368 [FREE Full text] [doi: 10.1093/aje/kwv011] [Medline: 25743324]
14. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015 Oct 07;13:167 [FREE Full text] [doi: 10.1186/s12955-015-0362-x] [Medline: 26446159]
15. Bennett AV, Dueck AC, Mitchell SA, Mendoza TR, Reeve BB, Atkinson TM, National Cancer Institute PRO-CTCAE Study Group. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Health Qual Life Outcomes 2016 Feb 19;14:24 [FREE Full text] [doi: 10.1186/s12955-016-0426-6] [Medline: 26892667]
16. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. J Clin Epidemiol 2014 Jan;67(1):108-113 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.07.016] [Medline: 24262772]
17. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE. Difference in method of administration did not significantly impact item response: an IRT-based analysis from the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative. Qual Life Res 2014 Feb;23(1):217-227 [FREE Full text] [doi: 10.1007/s11136-013-0451-4] [Medline: 23877585]
18. Kim JH, Kwon S, Shim SR, Sun HY, Ko YM, Chun D, et al. Validation and reliability of a smartphone application for the International Prostate Symptom Score questionnaire: a randomized repeated measures crossover study. J Med Internet Res 2014 Feb 10;16(2):e38 [FREE Full text] [doi: 10.2196/jmir.3042] [Medline: 24513507]
19. Ahmed R, Robinson R, Elsony A, Thomson R, Squire SB, Malmborg R, et al. A comparison of smartphone and paper data-collection tools in the Burden of Obstructive Lung Disease (BOLD) study in Gezira state, Sudan. PLoS One 2018;13(3):e0193917 [FREE Full text] [doi: 10.1371/journal.pone.0193917] [Medline: 29518132]
20. McManus DD, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Jung LS, et al. Design and Preliminary Findings From a New Electronic Cohort Embedded in the Framingham Heart Study. J Med Internet Res 2019 Mar 01;21(3):e12143 [FREE Full text] [doi: 10.2196/12143] [Medline: 30821691]
21. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol 2007 Jun 01;165(11):1328-1335. [doi: 10.1093/aje/kwm021] [Medline: 17372189]
22. Mohebbi M, Nguyen V, McNeil JJ, Woods RL, Nelson MR, Shah RC, ASPREE Investigator Group. Psychometric properties of a short form of the Center for Epidemiologic Studies Depression (CES-D-10) scale for screening depressive symptoms in healthy community dwelling older adults. Gen Hosp Psychiatry 2018;51:118-125 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.08.002] [Medline: 28890280]
23. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989 Mar;45(1):255-268. [Medline: 2720055]
24. Lin LI. Assay Validation Using the Concordance Correlation Coefficient. Biometrics 1992 Jun;48(2):599. [doi: 10.2307/2532314]
25. Redman S, Sanson-Fisher RW, Wilkinson C, Fahey PP, Gibberd RW. Agreement between two measures of alcohol consumption. J Stud Alcohol 1987 Mar;48(2):104-108. [doi: 10.15288/jsa.1987.48.104] [Medline: 3560945]
26. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]

27. Kannel WB, Belanger A, D'Agostino R, Israel I. Physical activity and physical demand on the job and risk of cardiovascular disease and death: the Framingham Study. Am Heart J 1986 Oct;112(4):820-825. [doi: 10.1016/0002-8703(86)90480-1] [Medline: 3766383]
28. Tan ZS, Spartano NL, Beiser AS, DeCarli C, Auerbach SH, Vasan RS, et al. Physical Activity, Brain Volume, and Dementia Risk: The Framingham Study. J Gerontol A Biol Sci Med Sci 2017 Jun 01;72(6):789-795 [FREE Full text] [doi: 10.1093/gerona/glw130] [Medline: 27422439]
29. Spartano N, Davis-Plourde K, Himali J, Murabito J, Vasan R, Beiser A, et al. Self-Reported Physical Activity and Relations to Growth and Neurotrophic Factors in Diabetes Mellitus: The Framingham Offspring Study. J Diabetes Res 2019;2019:2718465. [doi: 10.1155/2019/2718465] [Medline: 30729134]
30. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. J Youth Adolesc 1991 Apr;20(2):149-166. [doi: 10.1007/BF01537606] [Medline: 24265004]
31. Stahl D, Sum CF, Lum SS, Liow PH, Chan YH, Verma S, et al. Screening for depressive symptoms: validation of the center for epidemiologic studies depression scale (CES-D) in a multiethnic group of patients with diabetes in Singapore. Diabetes Care 2008 Jun;31(6):1118-1119. [doi: 10.2337/dc07-2019] [Medline: 18337303]
32. Sayetta R, Johnson D. Basic data on depressive symptomatology. United States, 1974-75. Vital Health Stat 11 1980 Apr;i-v(216):1-37 [FREE Full text] [doi: 10.1037/e531192008-001] [Medline: 7385607]
33. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Measure 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
34. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
35. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
36. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the Wild: Using Novel Data to Examine the Reach, Use, and Impact of PTSD Coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
37. Cohen J, Torous J. The Potential of Object-Relations Theory for Improving Engagement With Health Apps. JAMA 2019 Nov 14;322(22):2169. [doi: 10.1001/jama.2019.17141] [Medline: 31725854]
38. Zhen L, Wang G, Xu G, Xiao L, Feng L, Chen X, et al. Evaluation of the Paper and Smartphone Versions of the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR16) and the Patient Health Questionnaire-9 (PHQ-9) in Depressed Patients in China. Neuropsychiatr Dis Treat 2020;16:993-1001. [doi: 10.2147/NDT.S241766] [Medline: 32368061]
39. Šmigelskas K, Lukoševičiūtė J, Vaičiūnas T, Mozūraitytė K, Ivanavičiūtė U, Milevičiūtė I, et al. Measurement of Health and Social Behaviors in Schoolchildren: Randomized Study Comparing Paper Versus Electronic Mode. Zdr Varst 2019 Mar;58(1):1-10 [FREE Full text] [doi: 10.2478/sjph-2019-0001] [Medline: 30745945]
40. Njuguna HN, Caselton DL, Arunga GO, Emukule GO, Kinyanjui DK, Kalani RM, et al. A comparison of smartphones to paper-based questionnaires for routine influenza sentinel surveillance, Kenya, 2011-2012. BMC Med Inform Decis Mak 2014 Dec 24;14:107 [FREE Full text] [doi: 10.1186/s12911-014-0107-5] [Medline: 25539745]
41. Spartano NL, Lin H, Sun F, Lunetta KL, Trinquart L, Valentino M, et al. Comparison of On-Site Versus Remote Mobile Device Support in the Framingham Heart Study Using the Health eHeart Study for Digital Follow-up: Randomized Pilot Study Set Within an Observational Study Design. JMIR Mhealth Uhealth 2019 Sep 30;7(9):e13238 [FREE Full text] [doi: 10.2196/13238] [Medline: 31573928]
42. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
43. Badawy SM, Thompson AA, Kuhns LM. Medication Adherence and Technology-Based Interventions for Adolescents With Chronic Health Conditions: A Few Key Considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
44. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
45. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
46. Badawy SM, Thompson AA, Liem RI. Technology Access and Smartphone App Preferences for Medication Adherence in Adolescents and Young Adults With Sickle Cell Disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
47. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

48. Harrington CN, Wilcox L, Connelly K, Rogers W, Sanford J. Designing health and fitness apps with older adults: Examining the value of experience-based co-design. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, Pervasive Health 2018.: Association for Computing Machinery; 2018 May Presented at: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, Pervasive Health; 2018; New York p. 15-24 URL: https://dl.acm.org/doi/pdf/10.1145/3240925.3240929
49. Trinquart L, McManus DD, Nowak C, Kheterpal V, Manders ES, Fusco K, et al. Increasing Engagement in the Electronic Framingham Heart Study (eFHS): A Factorial Randomized Trial. In: Circulation.: American Heart Association; 2019 Presented at: Epidemiology, Big Data, and Precision Medicine; November 2019; Philadelphia p. A13084-A13084 URL: https://www.ahajournals.org/doi/10.1161/circ.140.suppl_1.13084
50. Jiang X, Ming W, You JH. The Cost-Effectiveness of Digital Health Interventions on the Management of Cardiovascular Diseases: Systematic Review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]
51. Jiang X, Yao J, You JH. Telemonitoring Versus Usual Care for Elderly Patients With Heart Failure Discharged From the Hospital in the United States: Cost-Effectiveness Analysis. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e17846 [FREE Full text] [doi: 10.2196/17846] [Medline: 32407288]

Abbreviations

aOR: adjusted odds ratio
CCC: concordance correlation coefficient
CES-D: Center for Epidemiologic-Studies Depression Scale
CVD: cardiovascular disease
eFHS: eCohort Framingham Heart Study
FHS: Framingham Heart Study
mHealth: mobile health
PAI: Physical Activity Index

Edited by G Eysenbach; submitted 14.10.20; peer-reviewed by S Badawy, Y Wu; comments to author 05.11.20; revised version received 15.12.20; accepted 19.12.20; published 20.01.21.

Please cite as:
Pathiravasan CH, Zhang Y, Trinquart L, Benjamin EJ, Borrelli B, McManus DD, Kheterpal V, Lin H, Sardana M, Hammond MM, Spartano NL, Dunn AL, Schramm E, Nowak C, Manders ES, Liu H, Kornej J, Liu C, Murabito JM
Adherence of Mobile App-Based Surveys and Comparison With Traditional Surveys: eCohort Study
J Med Internet Res 2021;23(1):e24773
URL: http://www.jmir.org/2021/1/e24773/
doi:10.2196/24773
PMID:33470944

Original Paper

Relationship Between Internet Use and Change in Health Status: Panel Study of Young Adults

Amanda Hunsaker[1], PhD; Eszter Hargittai[1], PhD; Marina Micheli[2], PhD

[1]Department of Communication and Media Research, University of Zurich, Zurich, Switzerland
[2]European Commission, Varese, Italy

Corresponding Author:
Eszter Hargittai, PhD
Department of Communication and Media Research
University of Zurich
Andreasstrasse 15
Zurich, 8050
Switzerland
Phone: 41 44 635 20 24
Email: pubs@webuse.org

Abstract

Background: Using the internet for health information is a widespread phenomenon documented in considerable scholarship. Less common, however, is the analysis of panel data to examine how internet use may relate to change in health status over time.

Objective: This study examines whether internet use and internet use related to health are associated with a change in health status among young adults.

Methods: We used a unique panel survey data set collected about young adults' internet use in 2012 and 2016 (n=384). We applied logistic regression to examine the relationships between sociodemographics, internet experiences, frequency of health-related internet use, and sharing health content online with change in health status over time. We additionally examined the variables characterizing sharing health content online (via Facebook, Twitter, and email) in separate models.

Results: In the second wave, over half (236/384, 61.5%) of the sample used the internet for health at least weekly. Approximately one-third (141/384, 36.7%) used Facebook for health-content sharing, while using Twitter and email for sharing health content were far less frequent (14/384, 3.6%, and 55/384, 14.3%, respectively). A change in health status occurred for 43.0% (165/384) of the sample; 18.5% (71/384) reported an improvement while 24.5% (94/384) reported a decline. Greater frequency of internet use was associated with health decline over time (*B*=–0.58, *P*=.02). We also found that frequent health-related internet use was related to enhanced health or maintained health (*B*=0.58, *P*=.03). Sharing health content on social media or email, however, was not related to young adults' health changes.

Conclusions: Young adults exhibit a pattern of using the internet for health that influences their health status. Our finding that frequent health-related internet use may promote improved or maintained health suggests that this type of online activity might also support healthy living.

(J Med Internet Res 2021;23(1):e22051) doi:10.2196/22051

KEYWORDS

internet use; health-related internet use; health status; panel data; young adults; internet; healthy living

Introduction

The internet has become an important means through which people engage with health content; among US adults, well over half turn to the internet as a source of health information [1-3]. Young adults—individuals in their late teens and twenties [4]—report even higher levels of internet use for health [5-7]. Young adults are over 2 times as likely to search for a health provider online and over 3 times as likely to search for health information as their older counterparts even when controlling for other sociodemographic characteristics [8]. Younger age is continually reported to be an indicator of a higher likelihood of using the internet for health-related reasons [9,10].

A substantial body of work has documented the relationship between internet use and social inequalities [11-14]. However, how internet use relates to the production of health inequalities or differences in health conditions and care due to social factors is less known [15]. Internet researchers continue to grapple with whether and how the proliferation of digital technology benefits or harms the well-being of individuals and society [16].

The internet offers several pathways to seek and share health-related information that may enhance medical knowledge. Examining the relationship between internet use and health may therefore inform whether the internet has the potential to influence the state of health in positive or negative ways. We address this question by analyzing unique panel data about a diverse group of young adults to see whether their internet uses relate to a change in their health status over time. Panel data are especially relevant to such questions as they sidestep the limitation of cross-sectional studies, which do not allow for conjectures about whether health status is a result of or a cause for the way people use the internet.

Most studies explore health-related internet use in the form of health information–seeking: purposeful searching for health information using reliable sources or those of unknown reliability for one's self or others [17,18]. Such use among young adults is strikingly high; 94% report looking for health information online while 76% have viewed or read about another individual's health experiences [7]. Although such internet use is prevalent, social inequalities persist. Mirroring findings across all age groups [19-22], among young adults, women and those with higher socioeconomic status are more likely to seek health information online than men and people from less privileged backgrounds [6,7,23-25]. Differing from reports on older age groups [22], studies of young adults report little variation across racial and ethnic groups with regard to health information–seeking online [6,7,24].

Compared with the considerable literature that has looked at how people seek health content online, fewer studies have examined the sharing of such content [26]. Similar to research on content-sharing more generally [27], studies vary widely in how they conceptualize and measure the domains of shared health content. Studies also differ by whether they specify social media platforms. Facebook comprises the bulk of research describing posting behavior [28-30], but research has also considered other platforms such as Twitter [31] and health-focused online communities [32]. While some studies do use secondary data sets that leave platform unspecified [10,33], survey questions may group sharing health content with reading health content under the overarching domain of user-generated and shared content [8], which makes it difficult to know what to conclude about sharing in particular in such contexts.

Social network sites attract growing numbers of the US population for reading or sharing about health-related topics, including managing health concerns [1,3]. This figure is higher among young adults, with half sharing such content in 2016 [34]. Age group comparisons over time also find that younger individuals share health-related content more often than older and middle-aged adults [10]. Being younger and female as well as having lower income and education all associate with using social network sites for health [3,8,10,35]. Overall, we conclude that social inequalities related to health information–seeking online also appear to be associated with sharing health content online. These findings, as well as the wealth of research documenting the link between social inequalities and health outcomes [36], set the stage for examining the relationship between health-related internet use and health outcomes.

Research exploring the relationship between internet use and self-reported general health status finds varying results. Some studies point to a positive association where better health relates to both general internet use [29,37,38] as well as more frequent use of the internet [39-41]. The relationship between health-related internet use specifically and health status is mixed and, adding further uncertainty, this variation in relationships occurs whether modeling health status as an independent variable [20,21,42-45] or outcome variable [41,46-48]. We review these streams of research in turn below.

Most research treats health status as an independent variable and models health information–seeking as the outcome of interest. Studies analyzing cross-sectional data from the Pew Research Center found that poor health was associated with health-related internet use [21,43]. Worse health was also related to a greater number of health topics sought online [21]. Among adults who looked for health information online, individuals reporting poorer health searched for health content online more frequently and completed specific searches for health care and treatment [43].

However, studies have also related better health to health information–seeking online. Cotten and Gupta [20] found that those who were healthier were more likely to seek health information online than via offline sources. The same significant relationship between better health and health-related internet use occurred in a multinational sample of adults, although certain indicators of worse health (eg, long-term illness and disability, health care use) did link to more health-related internet use [42]. A secondary analysis found that while health status was not associated with health information–seeking online, presence of a health condition did have a positive relationship [44]. The above findings reveal a clear variation across research studies, warranting further inquiry.

Far fewer studies examine health status as the outcome of interest relative to health-related internet use, yet all report no significant direct relationship between health-related internet use and health state [41,46-48]. Most such studies use cross-sectional samples [47,48], while panel data are much less common [41,46]. Research that looks at other health outcomes such as the presence of health-promoting behaviors does report significant associations with health information–seeking online [49].

Similar to health outcomes associated with health-related internet use, just a handful of studies specifically examine the relationship between health-related social network site use and health status, although none of these studies model health as an outcome variable [3,10,35,40]. Findings are mixed; one study found a positive relationship between having a primary care provider and social network site use for health [35]. Other studies reported no association between health status and health-related social network site use [3,10].

The potential benefits of social media for health communication include a range of positive implications for individuals in terms of social-emotional support, greater and more tailored health content, and increased social interaction [50]. That said, the frequency of time spent online may have a detrimental effect on well-being. A systematic review examining the impact of social media on well-being among adolescents reported overall mixed findings but noted that greater time spent online related to increased risk of exposure to harm [51]. Examining whether and by what mechanisms internet use influences health therefore becomes all the more important to tease out, especially for a group that spends considerable time online—young adults.

The purpose of this study is to examine the relationship between internet use and health change. We use panel data, which to our knowledge is used less often, to explore whether general internet use and internet use related to health are associated with a change in health status. We examine this relationship within a sample of young adults as this population is of particular relevance due to their high consumption of online health information [5,7]. We define the domains of health-related internet use as the frequency of internet use for health and sharing health content online, offering a more in-depth description of a little-explored area than most existing studies.

Methods

Data Collection

We used a unique panel survey data set collected about young adults' internet use in 2012 and 2016. The study received approval from the institutional review board of the principal investigator's home university (Northwestern University) at the time of data collection.

We surveyed students from a required first-year writing course at a US midwestern public university, thus avoiding any bias associated with who may be more likely to take specific classes. We administered the survey to students in 86 of the 92 course sections that agreed to participate in the study. Of the students enrolled in the class, 80.5% (1325/1646) completed the questionnaire. Of these respondents, we focused our study on first-year students under age 30 years; these 1115 people make up our initial 2009 sample. We then administered waves 2 and 3 through postal mail. Wave 2 (2012) occurred 3 years after the initial data collection, and respondents were offered a US $25 gift card for their participation. Wave 3 (2016) followed 4 years later with a US $40 gift card for those who returned their survey. In both years, we sent an introductory letter before the survey and then followed up after the survey with 2 postcard reminders and one more copy of the survey. In wave 2, the response rate was 51.2% (571/1115), but we discarded 24 cases (24/571, 4.2%) for not passing the attention check question, resulting in 547 valid responses. Wave 3 achieved a 34.44% (384/1115) response rate in comparison to the baseline sample, while 70.20% (384/547) of respondents completing the 2012 survey also completed the 2016 survey. Those completing the survey in waves 2 and 3 comprise our sample, and they mirror wave 1 except for a somewhat smaller proportion of African Americans. We used measures of internet use from 2016 (wave 3) data. We used measures of health status from 2012 and 2016 to capture change in health over 4 years. Sociodemographics come from the 2009 (baseline) survey.

Measures

Demographic Characteristics

We measured gender by asking participants whether they identify as male or female (see Multimedia Appendix 1 for all questions used in analyses). We collected additional data on age, race, and ethnicity separately. As a proxy for socioeconomic status, we asked respondents for the educational attainment of both their mother and father. We included the highest level of education completed by either parent in the analyses and recoded categories to high school or less, some college, and college degree or more.

Internet Experiences

As summarized in Table 1, we measured internet experiences in several ways: years of internet use, autonomy of use, frequency of use, and internet skills. We used data collected in 2016 for internet experiences with the exception of length of internet use which we measured in 2009. Measures of length of internet use were not repeated because years reported would increase proportionally in waves 2 and 3. We measured this by asking respondents how long they had been online. We measured weekly time spent online using categorical response options asking separately about weekdays and weekends, which we then recoded by multiplying the weekday response by 5 and the weekend response by 2, and then summing these for a possible range of 0 to 42. We logged this final number of hours to account for diminishing returns of additional time online in the regression models. To measure number of access locations, or autonomy of use, we asked respondents whether they use the internet from 10 different locations. As with number of hours online, we logged number of access points for logistic regression to address the diminishing returns of having greater numbers of internet access points. To gauge internet skills, participants completed a series of questions regarding their understanding of 27 internet-related terms [52]. Past research validated this measure with observational measures and used it with young adult samples [12,53].

Table 1. Sample characteristics (n=384).

Characteristics	Values
Gender, n (%)	
Women	231 (60.2)
Men	153 (39.8)
Race/ethnicity, n (%)	
White	167 (43.5)
African American	29 (7.6)
Hispanic	85 (22.8)
Asian	90 (23.4)
Parental education, n (%)	
High school or less	94 (24.5)
Some college	98 (25.5)
College degree or more	188 (49.0)
Missing	4 (1.0)
Internet experiences, mean (SD)	
Years of use[a] (in 2009)	5.6 (2.2)
Number of access locations	7.8 (1.7)
Weekly web hours	21.4 (10.5)
Internet skills (1-5)	3.6 (0.8)

[a] Years of use: n=382.

Health-Related Internet Use

We assessed two domains of health-related internet use in wave 3 (2016): frequency of overall health-related internet use and sharing health content online. We measured frequency of using the internet to obtain health information using a single question with categorical response options. We dichotomized this variable for regression modeling to monthly or less and weekly or more. To measure sharing health content we asked about such activity in the past year for Facebook, Twitter, and email, separately. We treated the measures of the use of each platform as separate, dichotomous measures in regression modeling.

Health Status

In 2012 and 2016 we measured health status with an extensively used, single-item question derived from the 36-Item Short Form Health Survey [54]. This measure allowed for brevity in questioning and is significantly related to other measures of health, including health care use, functional status, varied health conditions, and mortality [55-57]. We determined change in health over time by taking the 2016 health status measure and subtracting it from the 2012 measure. Positive values indicated improvement, zero indicated no change, and negative values indicated health decline. We recoded this variable into a binary measure of whether someone's health declined (=0) versus improved or stayed the same (=1).

Analysis

Given that none of the correlations of our measures of internet experiences and health-related internet use were higher than 0.2 (i.e., there was no multicollinearity in our data set), we continued with regression modeling. We used logistic regression to examine the relationships between sociodemographics, internet experiences, frequency of health-related internet use, and sharing health content online with change in health status (level of significance: $P<.05$). We examined the variables characterizing sharing health content online (via Facebook, Twitter, and email) in separate models.

Results

Sociodemographics

Table 1 summarizes the sociodemographic characteristics of respondents who completed surveys in waves 2 and 3 of data collection. There were more women than men in the sample (231/384, 60.2% versus 153/384, 39.8%), and their proportion is almost identical to the first wave of the study. The vast majority of respondents were aged 18 or 19 years at the time of the first data collection, and aged 25 to 26 years at the time of wave 3. Since age does not vary in the sample, we do not include it in analyses. Less than half (167/384, 43.5%) were White, almost one-quarter (90/384, 23.4%) Asian, a slightly smaller number (85/384, 22.8%) identified as Hispanic or Latinx, and less than 1 in 10 (29/384, 7.6%) as African American. The remaining respondents (13/384, 3.4%) reported being Native American, other, or did not report race/ethnicity. We found considerable variation in parental education; almost half (188/384, 49.0%) had at least one parent who completed at least a college degree, while roughly one-quarter (98/384, 25.5%)

had completed some college and a quarter (94/384, 24.5%) had no more than a high school degree.

Internet Experiences

Descriptions of internet experiences are found in Table 1 and show that while most participants have considerable online experience, the standard deviations signal that there is plenty of variation in the sample.

Health-Related Internet Use

Our results reveal both high prevalence and variation in internet use for health among young adults. Table 2 summarizes the frequency of such use. In 2016, over half (236/384, 61.5%) of the sample used the internet for health at least weekly, with percentages of the full sample roughly split between once a week, a few times a week, and daily. About a seventh (52/384, 13.5%) of the sample only did this a few times a year and just a handful (14/384, 3.6%) reported never using the internet for health. Sharing health content was a less common phenomenon. Facebook was the most common platform for health-content sharing with approximately one-third (141/384, 36.7%) of the sample doing so in 2016. Using Twitter and email for sharing health content were even less frequent.

Table 2. Characteristics of health-related internet use in 2016.

Characteristics	Values, n (%)
Frequency of health-related internet use (n=384)	
Never	14 (3.6)
Few times a year	52 (13.5)
Monthly	82 (21.4)
Weekly	85 (22.1)
Few times a week	82 (21.4)
Daily	69 (18.0)
Sharing health content online (n=383)	
On Facebook	141 (36.7)
On Twitter	14 (3.6)
On email	55 (14.3)

Change in Health Status

Table 3 describes the health status of the sample in 2012 and 2016. As expected for a young adult sample, in both years the majority of respondents reported either excellent or very good health. Although we found comparable descriptive findings across years, a notable number of young adults reported a differing health status from 2012 to 2016. A change in health status occurred for just under half (165/384, 43.0%) of the sample; about a fifth (71/384, 18.5%) reported an improvement (eg, fair to very good or very good to excellent) while a quarter (94/384, 24.5%) reported a decline (eg, excellent to very good or good to poor). Just over half (216/384, 56.3%) reported the same health status in 2012 and 2016. This is partly due to the fact that there is a ceiling effect. Over a third (136/384, 35.4%) of participants reported excellent health in 2012 meaning that these people could not have reported an improved health status in 2016 given our measure of health status. In fact, a fifth (77/384, 20.1%) of the sample remained in excellent health at the second time point. Given the young cohort, we did not have issues of floor effects with people reporting poor health in 2012 who then could not report worse than poor health in 2016. Just 2 respondents reported poor health in 2012, and both reported good health in 2016.

Table 3. Descriptives for health status and health status change (n=384).

Health status	2012, n (%)	2016, n (%)
Excellent	136 (35.4)	117 (30.5)
Very good	157 (40.9)	159 (41.4)
Good	73 (19.0)	91 (23.7)
Fair	15 (3.9)	11 (2.9)
Poor	2 (0.5)	4 (1.0)
Missing	1 (0.3)	2 (0.5)

Factors Related to Health Improving or Staying the Same

Using logistic regression, we modeled sociodemographic and internet-use factors related to health change, where we looked at health improving or staying the same compared with a decline in health (level of significance: *P*<.05). We completed modeling in which we first added sociodemographic characteristics (gender, race/ethnicity, parental education) to the model, and then internet experiences (autonomy of use, frequency of use, internet skills) followed by frequency of health-related internet use. Finally, we added the variables measuring sharing health content online through different platforms (Facebook, Twitter, and email) each in its own model. Table 4 presents these findings.

In the first model, we found no significant relationships between sociodemographics and health change. Turning to the second step, where we added the variables measuring internet experiences, being female emerged as significant and associated with positive health change/status quo health. Less time spent on the internet was significantly associated with health getting better or staying the same while holding sociodemographic factors constant. Additionally, having better internet skills significantly related to health improving or staying the same. Autonomy of use, or the number of access points for using the internet, was not statistically significant.

The third model included a dummy variable for regular health-related internet use (operationalized as weekly or more). None of the sociodemographics (gender, race/ethnicity, parental education) were significant in this model. Lower frequency of internet use continued to relate to improved or the same health status, while better internet skills no longer associated with better or status quo health. Autonomy of use stayed statistically nonsignificant in the model. Our newly added variable, health-related internet use at least weekly, related to health improving or staying the same when sociodemographics and internet experience variables were held constant.

Table 4. Logistic regression on health status improving or staying the same.

Variable	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	B [a]	*P* value	*B*	*P* value	*B*	*P* value	*B*	*P* value	*B*	*P* value	*B*	*P* value
Sociodemographics												
Gender, female	0.42	.09	0.58	.03	0.52	.06	0.52	.06	0.54	.05	0.50	.07
Race/ethnicity												
White	ref[b]	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
Hispanic	–0.06	.85	0.03	.92	–0.03	.93	–0.03	.94	–0.01	.99	–0.02	.95
African American	–0.71	.11	–0.56	.22	–0.62	.18	–0.62	.18	–0.62	.19	–0.62	.18
Asian	0.19	.56	0.24	.46	0.25	.46	0.25	.46	0.24	.47	0.25	.45
Parental education												
High school or less	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
Some college	–0.12	.75	–0.16	.67	–0.24	.52	–0.23	.53	–0.24	.52	–0.23	.54
College degree or more	–0.03	.93	–0.13	.72	–0.19	.58	–0.19	.59	–0.20	.57	–0.20	.57
Internet experiences												
Number of access locations	—	—	0.43	.33	0.26	.56	0.26	.56	0.29	.51	0.26	.56
Web use, hours per week	—	—	–0.58	.02	–0.62	.01	–0.62	.01	–0.63	.01	–0.62	.01
Internet skills	—	—	0.36	.03	0.32	.06	0.32	.06	0.37	.03	0.32	.06
Health-related internet use												
Health-related internet use weekly or more	—	—	—	—	0.58	.03	0.59	.03	0.63	.02	0.56	.04
On Facebook	—	—	—	—	—	—	–0.05	.86	—	—	—	—
On Twitter	—	—	—	—	—	—	—	—	–1.16	.06	—	—
Through email	—	—	—	—	—	—	—	—	—	—	0.21	.60
Constant	0.96	.01	0.49	.70	0.83	.51	0.83	.51	0.63	.62	0.83	.51

[a] B: unstandardized beta coefficient.

[b] ref: reference.

We next added sharing health content online using Facebook, Twitter, and email by separately creating three different models. As before, no sociodemographics (gender, race/ethnicity, and parental education) or autonomy of use associated with health change in any of the final modeling. However, less time spent online continued to relate to better or similarly good health as before, as did weekly use of the internet for health. Better internet skills again had a significant association with better or status quo health when modeled with using Twitter to share health information but had no association when sharing health content on Facebook or email. Said another way, when we included sharing health information via Twitter in modeling, having better internet skills significantly related to health improving or staying the same. Since only a tiny portion of respondents in the sample use Twitter for sharing health information, we hesitate to make too much of this finding. We found no significant association across these platforms for health-information sharing. We also ran these models without controlling for health-related internet use; sharing using the various means we examined remained nonsignificant (Table 5).

Table 5. Logistic regression on health status improving or staying the same modeling sharing health content online.

Variable	Model 1		Model 2		Model 3	
	B [a]	*P* value	*B*	*P* value	*B*	*P* value
Sociodemographics						
Gender, female	0.58	.03	0.60	.03	0.50	.07
Race/ethnicity						
White	ref[b]	ref	ref	ref	ref	ref
Hispanic	0.03	.94	0.06	.88	–0.02	.95
African American	–0.56	.22	–0.56	.23	–0.62	.18
Asian	0.24	.47	0.24	.47	0.25	.45
Parental education						
High school or less	ref	ref	ref	ref	ref	ref
Some college	–0.17	.65	–0.16	.66	–0.23	.54
College degree or more	–0.13	.71	–0.13	.71	–0.20	.57
Internet experiences						
Number of access locations	0.41	.34	0.47	.28	0.26	.56
Web use, hours per week	–0.58	.02	–0.59	.01	–0.62	.01
Internet skills	0.36	.03	0.41	.02	0.32	.06
Sharing health content online						
On Facebook	0.08	.77	—	—	—	—
On Twitter	—	—	–1.00	.10	—	—
Through email	—	—	—	—	0.21	.60
Constant	0.49	.69	0.29	.82	0.83	.51

[a] B: Unstandardized beta coefficient.

[b] ref: reference.

Discussion

Principal Findings

In this section, we reflect on our findings, discuss the limitations of our study, and suggest avenues for future research. In sum, having a decline in health significantly associated with greater overall time spent online and also less time explicitly using the internet for health. Sharing health content using different social network sites (Facebook or Twitter) or email did not relate to health change. We found that better internet skills related to positive health change or no health change but became nonsignificant when we included health-related internet use variables in the modeling, with the exception of adding health content shared on Twitter.

We found few previous studies that examined health over time in relation to internet use or health information seeking online [37,41,46,59] and the current findings offer a more in-depth look at the relationship between varying types and intensities of internet use and health change. In measuring health change over 4 years, our conceptualization of health improvement included young adults who reported an improvement from poorer health as well as those reporting no change in health. Young adults with no health problems may be accessing information to maintain their currently healthy state [60,61] and yet still gain a positive health change or maintain their health state. Studies examining the relationship between internet use and health in cross-sectional data have found that individuals reporting medical problems and those in a healthy state use the internet for health-related reasons [62].

The finding that greater frequency of internet use is associated with health decline differs from past research, where more frequent internet use associated with better health [40], including in a longitudinal study [37]. Although we recommend the interpretation of a causal relationship with caution, from one perspective having a decline in health may free up time to be online, as one may have health conditions that make it difficult to engage in other activities. While this interpretation may concern different health conditions in varying ways, previous research reports no difference in the frequency of internet use in relation to differing illness types [63]. Alternatively, spending too much time online may be a detriment to health status. The frequency of varying types of online activities may also matter. For example, a survey of first-year college students found that more hours on the internet shopping, doing research, and playing games significantly associated with more depressive symptoms, while more hours spent in online communication associated with decreased symptoms of depression [64]. However, the internet has changed considerably since that study so more work in this domain is needed.

In this paper, we found a clear association between health-related internet use and health change. Our finding that frequent health-related internet use may promote improved or maintained health suggests that this type of online activity might also support healthy living. Recent qualitative and quantitative research supports the notion that these topics are of interest to young adults [60,61]. Our findings differ from previous longitudinal research reporting no association between health-related internet use and change in general health [46]. Cross-sectional studies using young adult samples have also reported varied findings for the relationship between health-related internet use and health status [23,61,65]. One study found a positive relationship between health-promoting behaviors and health-related internet use [23], while another found no relationship between health-related internet use and health status [61]. Further adding to the diversity of study findings for this age group, other research found that young adults who used the internet for health information also reported more adverse health conditions [65]. This variation across study findings may be due to the large variety of health-related activities in which young adults engage online. One study reported that those searching for general health information online reported better health, while individuals searching for disease-specific information reported poorer health [39]. We recommend that future research include more details about the kinds of health topics that young adults explore online in relation to health outcomes.

We found no relationship between three different ways of sharing health content online (using Facebook, Twitter, and email) and health change. The low percentages of respondents using these platforms for health matters may account for our null findings, especially in light of our relatively small sample size. Although recent research notes a significant uptick in health-related internet use among young adults, social network sites were the least used sources and also considered the least reliable [61]. This distrust may hinder how young adults in turn share health content as they may be reluctant to post what they or their peers deem questionable information. Alternatively, we may see an increase in sharing health content online, as increased health information–seeking may create a wealth of online lay experts who have an expressed interest in sharing their health knowledge.

We found a significant association between being female and better or maintained health. Past studies show that male college students may engage in more risky health behaviors [66], yet they may also report better health than their female counterparts [67]. Differing health measures may indeed better explicate the relationship between health and varying ways of using the internet for health reasons as they relate to gender. Internet skills have often been associated with more capital-enhancing uses of the internet [12]—that is, uses from which they may benefit.

Most respondents reported their health as excellent or very good at both time points. Comparing these figures to national survey data of students enrolled in colleges or universities in the fall of 2012 [68] and 2016 [69] suggests that our sample reported better health overall at both time points and more respondents reported excellent health. However, unlike those studies, ours did not focus on health and thus did not prime respondents to think about health, which may explain the differences. It is also important to note that, compared with other studies that investigate the relationship of internet use and health, our study examined both as naturally occurring and is absent of an intervention design. Therefore, noting an association between these domains is all the more difficult to detect.

Health care professionals should understand the roles that health-related internet use plays for young adults, particularly if it has the potential to impact health in positive ways. Our results point to the potential importance of online tools for health that are specifically designed for young adults. To date, reviews find high-quality intervention studies specifically for this age group to be limited [70], although there is a growing body of studies using social network sites as platforms for health-behavior interventions for all ages [71]. How to harness such engagement among young adults for positive health behavior change warrants further exploration.

Limitations

Our findings should be considered with some care. We measured the natural occurrence of internet use and health through self-reports. There may be unobserved variables affecting our findings. Our small sample size, due to attrition typical of panel studies, as well as the low variation in health status and low prevalence of sharing health content online require our findings to be interpreted with caution. Additionally, because our study began when the internet did not offer the multitudes of options it does today, our measures of its uses for health purposes are not as detailed as current options would warrant. Future studies with larger samples sizes and greater variation across measures may further explore whether, in fact, sharing health content online relates to health outcomes. Despite these limitations, given that panel studies of internet use are rare, the paper nonetheless makes a unique contribution to understanding the longer-term consequences of internet use for health.

Conclusions

Analyzing a unique panel survey of a diverse group of young adults, this paper contributes to research on digital media by examining how internet use relates to change in health status over time. We find a relationship between internet use and longitudinal change in health, with frequent health-related internet use related to enhanced health or maintained health. We find no relationship of sharing health content on social media or email with health status. Given the relatively small size of our sample due to attrition in panel studies, future research may revisit that question on larger samples. Other measures of health outcomes, including specific health conditions and behaviors, as well as health care use, may also tease out the impact of internet use on health. Young adults exhibit a pattern of using the internet for health that influences their health status. Finding the avenues of internet use that may be most beneficial to their health and well-being is important to serving this generation just entering adulthood, as well as generations that follow.

Acknowledgments

The authors are grateful to the John D and Catherine T MacArthur Foundation and the Robert and Kaye Hiatt Fund at Northwestern University for their support of the data collection. They thank Sam Mandlsohn, Cierra Levy, Richard Herndon, and Devon Moore for their help with the 2016 data collection and the many research assistants of the Web Use Project in earlier years for their assistance. MM was affiliated with the Department of Communication and Media Research at the University of Zurich while working on this study and is currently affiliated with the European Commission.

Authors' Contributions

EH and AH conceived the research questions and analyzed the data. AH wrote the first draft of the manuscript. EH planned, oversaw, and completed the data collection and contributed to the writing of the manuscript. MM contributed to the writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.
[DOCX File , 27 KB - jmir_v23i1e22051_app1.docx]

References

1. 2018 Consumer Survey on Digital Health. Accenture. 2018. URL: https://www.accenture.com/t20180306T103559Z__w__/us-en/_acnmedia/PDF-71/accenture-health-2018-consumer-survey-digital-health.pdf [accessed 2020-11-12]
2. Hung L, Lyons JG, Wu C. Health information technology use among older adults in the United States, 2009-2018. Curr Med Res Opin 2020 May;36(5):789-797. [doi: 10.1080/03007995.2020.1734782] [Medline: 32096650]
3. Prestin A, Vieux SN, Chou WS. Is online health activity alive and well or flatlining? Findings from 10 years of the Health Information National Trends Survey. J Health Commun 2015;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
4. UN Department of Economic and Social Affairs—Youth: Frequently Asked Questions. United Nations. URL: https://www.un.org/development/desa/youth/what-we-do/faq.html [accessed 2020-11-12]
5. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet and American Life Project; 2013 Jan 15. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2020-11-12]
6. Percheski C, Hargittai E. Health information-seeking in the digital age. J Am Coll Health 2011;59(5):379-386. [doi: 10.1080/07448481.2010.513406] [Medline: 21500056]
7. Rideout V, Fox S. Digital health practices, social media use, and mental well-being among teens and young adults in the U.S. 2018. URL: https://assets.hopelab.org/wp-content/uploads/2020/08/a-national-survey-by-hopelab-and-well-being-trust-2018.pdf [accessed 2020-11-12]
8. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
9. Bach RL, Wenz A. Studying health-related internet and mobile device use using web logs and smartphone records. PLoS One 2020;15(6):e0234663 [FREE Full text] [doi: 10.1371/journal.pone.0234663] [Medline: 32530937]
10. Huo J, Desai R, Hong Y, Turner K, Mainous AG, Bian J. Use of social media in health communication: findings from the Health Information National Trends Survey 2013, 2014, and 2017. Cancer Control 2019;26(1):1-10 [FREE Full text] [doi: 10.1177/1073274819841442] [Medline: 30995864]
11. Bonfadelli H. The internet and knowledge gaps: a theoretical and empirical investigation. Eur J Commun 2002 Mar;17(1):65-84. [doi: 10.1177/0267323102017001607]

12. Hargittai E, Hinnant A. Digital inequality: differences in young adults' use of the internet. Commun Res 2008 Aug 04;35(5):602-621. [doi: 10.1177/0093650208321782]
13. Livingstone S, Helsper E. Balancing opportunities and risks in teenagers’ use of the internet: the role of online skills and internet self-efficacy. New Media Soc 2009 Nov 24;12(2):309-329. [doi: 10.1177/1461444809342697]
14. van Deursen AJ, van Dijk JA, Peters O. Rethinking Internet skills: the contribution of gender, age, education, internet experience, and hours online to medium- and content-related internet skills. Poetics 2011 Apr;39(2):125-144. [doi: 10.1016/j.poetic.2011.02.001]
15. Latulippe K, Hamel C, Giroux D. Social health inequalities and ehealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
16. Anderson J, Rainie L. The future of well-being in a tech-saturated world. Washington: Pew Internet and American Life Project; 2018 Apr 17. URL: https://assets.pewresearch.org/wp-content/uploads/sites/14/2018/04/14154552/PI_2018.04.17_Future-of-Well-Being_FINAL.pdf [accessed 2020-11-12]
17. Cline R, Haynes K. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001;16(6):671-692. [doi: 10.1093/her/16.6.671]
18. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
19. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
20. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
21. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2015 Aug 10;22(4):804-814. [doi: 10.1177/1460458215595851] [Medline: 26261218]
22. Mitchell U, Chebli P, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
23. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
24. Dobransky K, Hargittai E. Inquiring minds acquiring wellness: uses of online and offline sources for health information. Health Commun 2012;27(4):331-343. [doi: 10.1080/10410236.2011.585451] [Medline: 21932982]
25. Hassan S, Masoud O. Online health information seeking and health literacy among non-medical college students: gender differences. J Public Health (Berl.) 2020 Mar 9:1-7. [doi: 10.1007/s10389-020-01243-w]
26. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
27. Hargittai E, Jennrich K. The online participation divide. In: Lloyd M, Friedland LA, editors. The Communication Crisis in America, And How to Fix It. London: Palgrave Macmillan; 2016:199-213.
28. Bender JL, Jimenez-Marroquin M, Jadad AR. Seeking support on Facebook: a content analysis of breast cancer groups. J Med Internet Res 2011 Feb 04;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
29. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. J Med Internet Res 2014 Aug 04;16(8):e182 [FREE Full text] [doi: 10.2196/jmir.3275] [Medline: 25092386]
30. Lawless M, Augoustinos M, LeCouteur A. Dementia on Facebook: requesting information and advice about dementia risk-prevention on social media. Discourse Context Media 2018 Oct;25:44-51. [doi: 10.1016/j.dcm.2018.01.011]
31. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
32. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
33. Jackson DN, Chou WS, Coa KI, Oh A, Hesse B. Implications of social media use on health information technology engagement: data from HINTS 4 cycle 3. Transl Behav Med 2016 Dec;6(4):566-576 [FREE Full text] [doi: 10.1007/s13142-016-0437-1] [Medline: 27628436]
34. Hargittai E, Füchslin T, Schäfer MS. How do young adults engage with science and research on social media? Some preliminary findings and an agenda for future research. Soc Media Soc 2018 Sep 03;4(3):1-10. [doi: 10.1177/2056305118797720]
35. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013 Jan 30;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
36. Marmot M, Friel S, Bell R, Houweling TA, Taylor S, Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. Lancet 2008 Nov 08;372(9650):1661-1669. [doi: 10.1016/S0140-6736(08)61690-6] [Medline: 18994664]

37. Hartanto A, Yong JC, Toh WX, Lee ST, Tng GY, Tov W. Cognitive, social, emotional, and subjective health benefits of computer use in adults: a 9-year longitudinal study from the Midlife in the United States (MIDUS). Comput Hum Behav 2020 Mar:104-109. [doi: 10.1016/j.chb.2019.106179]
38. Wangberg S, Andreassen H, Kummervold P, Wynn R, Sørensen T. Use of the internet for health purposes: trends in Norway 2000-2010. Scand J Caring Sci 2009 Dec;23(4):691-696. [doi: 10.1111/j.1471-6712.2008.00662.x] [Medline: 19473317]
39. Hone T, Palladino R, Filippidis FT. Association of searching for health-related information online with self-rated health in the European Union. Eur J Public Health 2016 Oct;26(5):748-753. [doi: 10.1093/eurpub/ckw022] [Medline: 26944894]
40. Neter E, Brainin E, Baron-Epel O. Chapter 6: The third digital divide in the health domain: Is internet use for health purposes associated with health benefits? In: EHealth: Current Evidence, Promises, Perils and Future Directions. Vol 15. Studies in Media and Communications. Bingley: Emerald Publishing Limited; 2018:153-175.
41. Philbin MM, Parish C, Pereyra M, Feaster DJ, Cohen M, Wingood G, et al. Health disparities and the digital divide: the relationship between communication inequalities and quality of life among women in a nationwide prospective cohort study in the United States. J Health Commun 2019;24(4):405-412 [FREE Full text] [doi: 10.1080/10810730.2019.1630524] [Medline: 31198091]
42. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
43. Houston TK, Allison JJ. Users of Internet health information: differences by health status. J Med Internet Res 2002 Nov;4(2):e7 [FREE Full text] [doi: 10.2196/jmir.4.2.e7] [Medline: 12554554]
44. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
45. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
46. Bessière K, Pressman S, Kiesler S, Kraut R. Effects of internet use on health and depression: a longitudinal study. J Med Internet Res 2010 Feb 28;12(1):e6 [FREE Full text] [doi: 10.2196/jmir.1149] [Medline: 20228047]
47. Jiang S, Street RL. Pathway linking internet health information seeking to better health: a moderated mediation study. Health Commun 2017 Dec;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
48. Lee HY, Kim J, Sharratt M. Technology use and its association with health and depressive symptoms in older cancer survivors. Qual Life Res 2018 Feb;27(2):467-477. [doi: 10.1007/s11136-017-1734-y] [Medline: 29128998]
49. Chae J. How we use the Internet matters for health: the relationship between various online health-related activities and preventive dietary behaviors. Health Informatics J 2019 Sep;25(3):973-983 [FREE Full text] [doi: 10.1177/1460458217735675] [Medline: 29047292]
50. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
51. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: a systematic narrative review. Child Youth Serv Rev 2014 Jun;41:27-36. [doi: 10.1016/j.childyouth.2014.03.001]
52. Hargittai E, Hsieh YP. Succinct survey measures of web-use skills. Soc Sci Comput Rev 2011 Feb 28;30(1):95-107. [doi: 10.1177/0894439310397146]
53. Hargittai E. Survey measures of web-oriented digital literacy. Soc Sci Comput Rev 2005 Aug 01;23(3):371-379. [doi: 10.1177/0894439305275911]
54. Ware JE, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) project. J Clin Epidemiol 1998 Nov;51(11):903-912. [doi: 10.1016/s0895-4356(98)00081-x] [Medline: 9817107]
55. Bowling A. Just one question: if one question works, why ask several? J Epidemiol Community Health 2005 May;59(5):342-345 [FREE Full text] [doi: 10.1136/jech.2004.021204] [Medline: 15831678]
56. DeSalvo K, Fan V, McDonell M, Fihn S. Predicting mortality and healthcare utilization with a single question. Health Serv Res 2005 Aug;40(4):1234-1246 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00404.x] [Medline: 16033502]
57. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Soc Sci Med 2009 Aug;69(3):307-316. [doi: 10.1016/j.socscimed.2009.05.013] [Medline: 19520474]
58. Hargittai E. Potential biases in big data: omitted voices on social media. Soc Sci Comput Rev 2020;38(1):10-24. [doi: 10.1177/0894439318788322]
59. Hale T, Goldner M, Stern M, Drentea P, Cotten S. Patterns of online health searching 2002-2010: implications for social capital, health disparities and the de-professionalization of medical knowledge. In: Technology, Communication, Disparities and Government Options in Health and Health Care Services (Research in the Sociology of Health Care), Volume 32. Bingley: Emerald Group Publishing Limited; 2014:35-60.
60. Mendes Á, Abreu L, Vilar-Correia MR, Borlido-Santos J. “That should be left to doctors, that's what they are there for!” Exploring the reflexivity and trust of young adults when seeking health information. Health Commun 2017 Sep;32(9):1076-1081. [doi: 10.1080/10410236.2016.1199081] [Medline: 27485762]

61. Montagni I, Cariou T, Feuillet T, Langlois E, Tzourio C. Exploring digital health use and opinions of university students: field survey study. JMIR Mhealth Uhealth 2018 Mar 15;6(3):e65 [FREE Full text] [doi: 10.2196/mhealth.9131] [Medline: 29549071]
62. Haviland MG, Pincus HA, Dial TH. Datapoints: type of illness and use of the internet for health information. Psychiatr Serv 2003 Sep;54(9):1198-1198. [doi: 10.1176/appi.ps.54.9.1198] [Medline: 12954932]
63. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
64. Morgan C, Cotten SR. The relationship between internet activities and depressive symptoms in a sample of college freshmen. Cyberpsychol Behav 2003 Apr;6(2):133-142. [doi: 10.1089/109493103321640329] [Medline: 12804025]
65. Rowlands IJ, Loxton D, Dobson A, Mishra GD. Seeking health information online: association with young Australian women's physical, mental, and reproductive health. J Med Internet Res 2015 May 18;17(5):e120 [FREE Full text] [doi: 10.2196/jmir.4048] [Medline: 25986630]
66. von Bothmer MIK, Fridlund B. Gender differences in health habits and in motivation for a healthy lifestyle among Swedish university students. Nurs Health Sci 2005 Jun;7(2):107-118. [doi: 10.1111/j.1442-2018.2005.00227.x] [Medline: 15877687]
67. Dawson KA, Schneider MA, Fletcher PC, Bryden PJ. Examining gender differences in the health behaviors of Canadian university students. J R Soc Promot Health 2016 Sep 07;127(1):38-44. [doi: 10.1177/1466424007073205] [Medline: 17319316]
68. American College Health Association-National College Health Assessment II: Reference Group Data Report Fall 2012. American College Health Association. 2013. URL: https://www.acha.org/documents/ncha/ACHA-NCHA-II_ReferenceGroup_DataReport_Fall2012.pdf
69. American College Health Association-National College Health Assessment II: Undergraduate Student Reference Group Data Report Fall 2016. American College Health Association. 2017. URL: https://www.acha.org/documents/ncha/NCHA-II_FALL_2016_UNDERGRADUATE_REFERENCE_GROUP_DATA_REPORT.pdf [accessed 2020-11-12]
70. Hayba N, Partridge SR, Nour MM, Grech A, Allman Farinelli M. Effectiveness of lifestyle interventions for preventing harmful weight gain among young adults from lower socioeconomic status and ethnically diverse backgrounds: a systematic review. Obes Rev 2018 Mar;19(3):333-346. [doi: 10.1111/obr.12641] [Medline: 29178423]
71. Yang Q. Are social networking sites making health behavior change interventions more effective? A meta-analytic review. J Health Commun 2017 Mar;22(3):223-233. [doi: 10.1080/10810730.2016.1271065] [Medline: 28248623]

Edited by G Eysenbach; submitted 03.07.20; peer-reviewed by L Brennan, A Davies; comments to author 09.10.20; revised version received 20.10.20; accepted 28.10.20; published 13.01.21.

Please cite as:
Hunsaker A, Hargittai E, Micheli M
Relationship Between Internet Use and Change in Health Status: Panel Study of Young Adults
J Med Internet Res 2021;23(1):e22051
URL: https://www.jmir.org/2021/1/e22051
doi:10.2196/22051
PMID:33439134

Original Paper

Use of eHealth by Patients With Rheumatoid Arthritis: Observational, Cross-sectional, Multicenter Study

Marion Magnol[1], MD; Berard Eleonore[2], MD; Rempenault Claire[3], MD; Benjamin Castagne[4], MD; Marine Pugibet[5], MD; Cédric Lukas[3], MD, PhD; Anne Tournadre[6], MD, PhD; Pascale Vergne-Salle[5], MD, PhD; Thomas Barnetche[2], PhD; Marie-Elise Truchetet[2], MD, PhD; Adeline Ruyssen-Witrand[1], MD, PhD

[1]Rheumatology Department, Toulouse University Hospital, Toulouse, France
[2]Rheumatology Department, Bordeaux University Hospital, Bordeaux, France
[3]Rheumatology Department, Montpellier University Hospital, Montpellier, France
[4]Rheumatology Department, Le-Puy-En-Velay Hospital, Le-Puy-En-Velay, France
[5]Rheumatology Department, Limoges University Hospital, Limoges, France
[6]Rheumatology Department, Clermont-Ferrand University Hospital, Clermont-Ferrand, France

Corresponding Author:
Marion Magnol, MD
Rheumatology Department
Toulouse University Hospital
Centre de Rhumatologie Pôle inflammation, infection, immunologie et loco-moteur Hôpital Pierre-Paul Riquet Place du Docteur Baylac
TSA 40031
Toulouse, 31059 cedex9
France
Phone: 33 561776976
Email: marion.magnol13@gmail.com

Abstract

Background: The use of eHealth tools (eg, the internet, mobile apps, and connected devices) in the management of chronic diseases and for rheumatoid arthritis is growing. eHealth may improve the overall quality of care provided to patients with chronic diseases.

Objective: The primary objective of this study was to describe eHealth use by patients with rheumatoid arthritis in France. The secondary objectives were to identify associations between patient demographics and disease characteristics and the use of eHealth tools, and assess their expectations of eHealth.

Methods: In this cross-sectional, multicenter study, patients with rheumatoid arthritis, according to the 2010 ACR/EULAR classification criteria, were recruited from 5 university hospitals (Bordeaux, Clermont-Ferrand, Limoges, Montpellier, and Toulouse). Patients completed an anonymous self-questionnaire, including demographic data, evaluating their eHealth use (ie, access, support, frequency of use, type of use, and reason for use). The rheumatologist in charge of each patient completed an independent medical questionnaire on disease characteristics, activity of rheumatoid arthritis, and treatments. Data were collected between December 2018 and July 2019.

Results: Questionnaires were completed by 575 participants, with a mean age of 62 (SD 13) years, 447 (77.7%) of whom were female. Overall, 82.2% (473/575) of the participants had access to eHealth through a computer (402/467, 86.1%), tablet (188/467, 40.2%), or smartphone (221/467, 47.3%). Of these, 36.4% (170/467) of the participants used the internet for health in general, and 28.7% (134/467) used it specifically for rheumatoid arthritis–related reasons. All these 134 patients used eHealth to learn about disease pathology, and 66.4% (89/134) of them used it as a tool to help monitor rheumatoid arthritis. Most patients (87/125, 69.6%) had a paper file, 19.2% (24/125) used a digital tool (spreadsheets, 10/125, 8%; mobile app, 9/125, 7.2%; or website, 5/125, 4%), and 24.8% (31/125) did not use any tools for monitoring. Few patients (16/125, 12.8%) used tools for treatment reminders. About 21.6% (27/125) of the patients using eHealth used a specific app for rheumatoid arthritis. Univariate analysis showed that age, education level, employment status, treatment, comorbidities, membership of a patient association, and patient education program were associated with eHealth use for rheumatoid arthritis. Multivariate analysis showed that membership of a patient association (odds ratio [OR] 5.8, 95% CI 3.0-11.2), use of biologic disease-modifying antirheumatic drugs (OR 0.6, 95% CI 0.4-1.0), and comorbidities (OR 0.7, 95% CI 0.6-0.8) remained associated with eHealth use for rheumatoid arthritis. Recommendation by a doctor (225/330, 68.2%), ease of use (105/330, 31.8%), and data security (69/330, 20.9%) were factors favoring the use of eHealth.

Conclusions: To date, few patients have used eHealth for disease management. The use of a reliable and validated eHealth tool for rheumatoid arthritis could therefore be promoted by rheumatologists and could optimize therapeutic adherence.

(J Med Internet Res 2021;23(1):e19998) doi:10.2196/19998

KEYWORDS

eHealth; internet; mobile app; rheumatoid arthritis; patients’ expectation

Introduction

Rheumatoid arthritis (RA) is one of the most common chronic autoimmune inflammatory rheumatic diseases with a prevalence of 0.1%-1% [1]. The disease is primarily characterized by a history of painful and swollen joints leading to joint deformation and destruction and disability. Nowadays, a wide range of disease-modifying antirheumatic drugs (DMARDs) are available (eg, synthetic, biologic, and targeted synthetic DMARDs). Because RA is a chronic disease, DMARDs are long-term maintenance treatments used to control the activity of rheumatism and prevent further joint destruction and disability.

In other chronic diseases such as diabetes, asthma, and hypertension, the use of eHealth has been well established. eHealth could be defined as an overreaching term used to describe the application of information and communication technologies in the health sector [2]. Studies show that eHealth and eHealth tools (ie, mobile apps, internet-based software and websites, connected devices, and personal health records) are ways to enhance medication adherence and disease control [3-5]. These eHealth tools provide information about the disease and help patients to monitor and manage the disease by improving patient autonomy.

Several websites, mobile apps, and connected devices related to RA have been developed for patients to obtain information about, self-monitor, or self-manage the disease. However, the impact of these specific eHealth services on disease management, medication adherence, or quality of life has been poorly assessed in the literature [6-9].

To date, there are no studies investigating the use of eHealth tools by patients with RA in France. Therefore, the main objective of this study was to describe the use of eHealth by patients with RA in France. The secondary objectives were to identify associations between patient demographics and disease characteristics and the use of eHealth tools, as well as to assess patients’ expectations of eHealth.

Methods

Study Design

A cross-sectional, multicenter, observational study was conducted in the rheumatology departments of 5 university hospitals in France (located at Bordeaux, Clermont-Ferrand, Limoges, Montpellier, and Toulouse). Data were collected from December 2018 to July 2019.

Study Patients

All adult outpatients and hospitalized patients were eligible to be included in this study if they were diagnosed with RA according to the 2019 ACR/EULAR (American College of Rheumatology/European League Against Rheumatism) classification criteria [10]. Eligible patients were systematically asked to participate in the survey, and if they agreed, they were included in the study. Patients who refused to participate or had language difficulties were excluded from the study. All patients provided written informed consent and agreed to participation. The local ethics committee (number 2018-A01875-50) approved this study.

Data Collection

Data were collected during a single visit through an anonymous patient self-questionnaire; this included sociodemographic data, medical data, and assessment of the use of eHealth tools (ie, access, support, frequency of use, type of use, and reason for use). The use of eHealth for RA was defined as “to possess an electronic tool” and “to use it to get information or to manage RA.” The rheumatologist in charge of the patient completed an independent medical questionnaire collecting RA characteristics and comorbidities (details on data collected are available in Multimedia Appendix 1).

Statistical Analysis

Qualitative variables were described as absolute values and percentages. Quantitative variables were described as mean (SD) or median (IQR) values. Demographic characteristics, disease characteristics, disease activity, treatment intake, and comorbidity index were compared between eHealth tools users and nonusers. Additionally, chi-square test or Fisher exact test, as required, was used to compare qualitative variables. Student *t*-test was used to compare quantitative variables with normal distribution and homogeneous variances. Logistic regression, with backward procedure to select variables, was used for multivariate analysis to identify patients’ characteristics that were independently associated with eHealth use for RA. All variables associated with eHealth use for RA with $P<.1$ in univariate analyses were tested in the model. The results are described with odds ratio (OR). Statistical analyses were carried out using STATA software (version 13.1; StataCorp).

Results

eHealth Use in RA populations

A total of 575 patients completed the self-questionnaire. We found that 82.2% (473/575) of the patients had an eHealth tool, and 28.7% (134/467) of them used it for RA. All of these 134 patients (100%) used the eHealth tool to obtain information about RA, and 89 (66.4%) of them used it to monitor their rheumatism. Figure 1 shows the use of eHealth tools in the study population. Table 1 shows the modalities of use and the frequency of access to eHealth for RA. Most patients used the internet infrequently.

Figure 1. Use of eHealth tools in the study population.

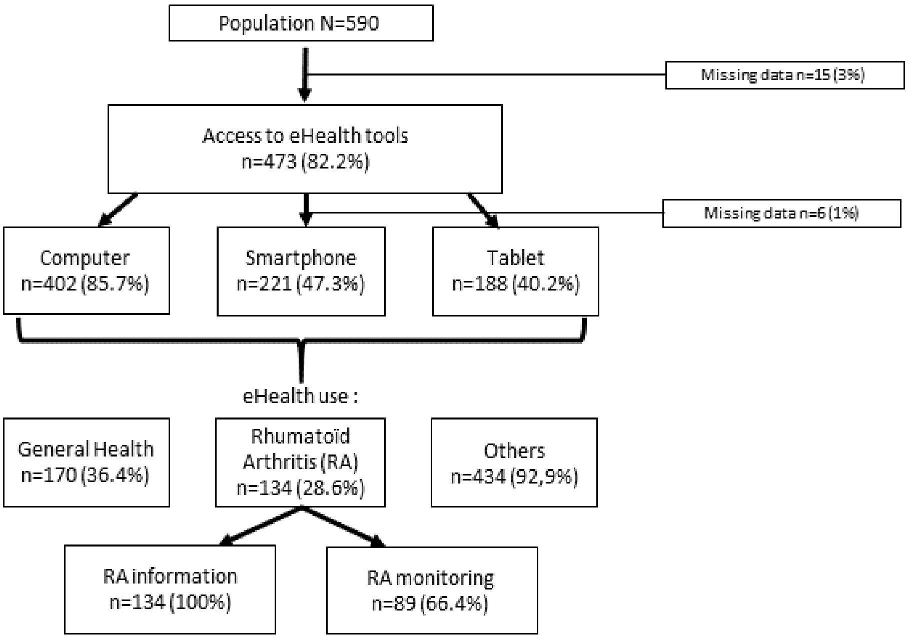


Table 1. Frequency and modalities of use of the internet by study patients using eHealth for rheumatoid arthritis.

Modality of use	Frequency of use		
	Often used (>1/month), n (%)	Sometimes used (≤1/month), n (%)	Never used, n (%)
eHealth content			
Rheumatoid arthritis information (n=128)	15 (11.7)	92 (71.8)	21 (16.4)
Treatment information (n=128)	21 (16.4)	83 (64.8)	24 (18.7)
Patient forums (n=122)	13 (10.6)	45 (36.8)	64 (52.4)
Social network (n=123)	16 (13)	16 (13)	91 (73.9)
Communication with the rheumatologist (n=123)	4 (3.2)	26 (21.1)	93 (75.6)
Appointment management (n=122)	8 (6.5)	33 (27)	81 (66.3)
Websites			
Generic French websites (eg, Wikipedia), (n=121)	16 (13.2)	55 (45.4)	50 (40.9)
Rheumatology-specific French websites (eg, rhumato.net, ANDAR, SPILF), (n=123)	28 (22.7)	55 (44.7)	40 (32.5)
Hospital center websites (n=119)	4 (3.3)	29 (24.3)	86 (72.2)
Pharmaceutical industry websites (n=120)	3 (2.5)	17 (14.1)	100 (83.3)

Association Between Patient Characteristics and eHealth Use for RA

Factors such as age, education level, employment, DMARDs, level of comorbidity, membership of a patient association, and participation in a patient education program were found to be associated with the use of eHealth tools for RA (Table 2).

Multivariate analysis showed that membership of a patient association remained independently associated with the use of eHealth tools for RA (OR 5.8, 95% CI 3.0-11.2; *P*<.001), whereas a high level of comorbidity (OR 0.7, 95% CI 0.6-0.8; *P*<.001) and use of biologic DMARDs (OR 0.6, 95% CI 0.4-1.0; *P*<.041) were associated with a lower use of eHealth tools.

Table 2. Patient characteristics and factors associated with the use of eHealth tools. Italicized values indicate statistically significant values.

Characteristic	Patients using eHealth for RA[a] (n=134)	Patients not using eHealth (n=441)	*P* value
Demographic characteristics			
Gender (female), n (%)	110/134 (82)	337/440 (76.5)	.18
Age (years), mean (SD), (n=569)	58.4 (13.2)	62.8 (12.1)	*<.001*
BMI (kg/m²), mean (SD), (n=553)	25.2 (5.0)	25.7 (5.1)	.32
Living place, n (%)			.35
Urban	60/132 (45.4)	174/426 (40.8)	
Rural	72/132 (54.5)	252/426 (59.1)	
Education level, n (%)			*.006*
Undergraduate or middle school	53/126 (42)	237/424 (55.8)	
High school or college	73/126 (57.9)	187/424 (43.3)	
Employment status, n (%)			*.006*
Farmer, artisan, worker	10/132 (7.5)	45/431 (10.4)	
Senior framework, employed	49/132 (37.1)	97/431 (22.5)	
Unemployed	19/132 (14.3)	58/431 (13.4)	
Retired	54/132 (40.9)	231/431 (53.5)	
RA characteristics			
RA duration (years), mean (SD), (n=551)	15.2 (12.0)	15.8 (10.0)	.56
RF[b]+, n (%)	101/123 (82.1)	333/421 (79.1)	.46
ACPA[c]+, n (%)	101/123 (82.1)	339/416 (81.4)	.88
Erosive, n (%)	73/125 (58.4)	274/423 (64.7)	.19
DAS28[d]-CRP[e], mean (SD), (n=505)	2.5 (1.2)	2.6 (1.2)	.84
Treatments			
Corticosteroids, n (%)	32/128 (25)	92/426 (21.5)	.42
csDMARDs[f] monotherapy (MTX[g] or other csDMARDs), n (%)	44/128 (34.3)	108/425 (25.4)	*.046*
bDMARDs[h] (IV[i], SC[j]), n (%)	75/128 (58.5)	304/425 (71.5)	*.006*
Association of csDMARDs + bDMARDs, n (%)	49/128 (38.2)	199/425 (46.8)	.09
Comorbidity (Charlson index), mean (SD), (n=545)	1.8 (1.5)	2.7 (2.0)	*<.001*
Membership of a patient association, n (%)	29/132 (21.9)	22/432 (5)	*<.001*
Patient education program, n (%)	47/131 (35.8)	88/428 (20.5)	*<.001*

[a] RA: rheumatoid arthritis.

[b] RF: rheumatoid factor.

[c] ACPA: anticitrullinated protein antibodies.

[d] DAS28: Disease Activity Score-28.

[e] CRP: c-reactive protein.

[f] csDMARDs: conventional synthetic disease-modifying antirheumatic drugs.

[g] MTX: methotrexate.

[h] bDMARDs: biologic disease-modifying antirheumatic drugs.

[i] IV: intravenous.

[j] SC: subcutaneous.

Regarding health applications, 21 (16.5%) and 27 (21.1%) of the 127 patients used ehealth applications (eg, sports, nutrition, diabetes, or cardiovascular) and specific rheumatic applications (Hiboot, J’agis, Arthritis, Sanoia, or myPR), respectively.

A limited group of patients used electronic technology for RA-related follow-up (Excel spreadsheets, 10/125, 8%; mobile app, 9/125, 7.2%; website, 5/125, 4%; paper record, 87/125, 69.6%; and none, 31/125, 24.8%) and medication reminder (electronic diary, 13/125, 10.4%; mobile app, 3/125, 2.4%; clock, 7/125, 5.6%; paper diary, 16/125, 12.8%), whereas 67.2% (84/125) of them used it for no specific reason.

Patients’ Expectations About eHealth for RA

When investigating patients’ expectations, 225 of 330 patients (68.2%) reported they would use eHealth to manage their RA if it were recommended by a doctor. eHealth device characteristics such as ease of use and security were mentioned by 105 (31.8%) and 69 (20.9%) of the 330 patients, respectively, to increase their adherence to eHealth, and 89 (27%) patients declared they would more likely use eHealth in the case of an RA flare.

Discussion

To the best of our knowledge, no study thus far has investigated the frequency of eHealth use in a specific population of RA. In this study, we found that 473 of 575 (82.2%) patients had electronic devices and about one-third of those used eHealth specifically for RA to obtain information about the disease and to help with disease management. However, specific use of a mobile app for RA as digital tool to remind themselves about medications or follow-up was rarely reported.

Digital device possession in this RA population was comparable to that in a French population in a previous study [11] and in other studies examining chronic diseases (eg, cancer and cardiovascular diseases), but it was lower than that reported in a diabetes population (55%-84% in available studies) [12-14] perhaps due to the more recent availability of these devices. Moreover, the therapeutic target is not objectively measurable by patients with RA, whereas blood glucose levels can be self-measured on a daily basis by patients with diabetes and consequent therapeutic changes can be decided by the patient after they have completed an educational program. Therapeutic changes for patients with RA, however, are complex decisions made by a physician.

As previously shown in studies examining other chronic diseases [12,14-16], membership of a patient association is strongly associated with eHealth use, which suggests that networking among patients is an effective way to enhance eHealth use. Closer follow-up (eg, day hospitalizations and consultations) of patients with RA receiving biotherapy may explain the reduced need for eHealth. Similarly, patients with increased comorbidities, who are also older, are more likely to make more frequent medical visits and use eHealth less frequently.

The low frequency of eHealth use for RA could be explained by patients and disease barriers. For example, older patients use eHealth less often, and pain and joint deformations can restrict the use of eHealth devices among these patients. Patients’ habits and disease activity may also affect the usefulness of eHealth (eg, patients in sustained remission might regard the use of eHealth to monitor their disease as futile) [6,17,18].

Other limitations include the lack of information provided by health professionals about the different eHealth tools available and the discrepancy of eHealth use by physicians [19], thus resulting in the lack of promotion of these instruments. Finally, evidence-based medicine [20] and data security [21] are fundamental aspects; however, the utility and data security of the majority of eHealth tools have never been assessed.

Our study has several limitations. First, data were collected with a declarative questionnaire that may not reflect the exact use of eHealth and increase the risk of missing data or false answers. However, the rate of missing data was very low in this study (ie, 3%), and false answers were minimized since the questionnaires were anonymously completed by patients who were informed that their rheumatologists would remain blinded to their answers. The strength of this study was its multicenter design with a systematic inclusion of consecutive patients, which resulted in a representative sample of patients with RA in France.

In the context of the current COVID-19 pandemic where in-person consultations are limited, eHealth, which allows for remote monitoring, remote auto-evaluation of disease activity, and teleconsultation, has gained considerable interest and continues to be developed [22].

In summary, the frequency of eHealth use for RA is low in France, especially in patients with multimorbidities and severe disease and those using biologic DMARDs. Further studies need to be conducted to evaluate the reasons regarding the use and nonuse of eHealth by patients with RA. Furthermore, studies assessing the efficacy of eHealth tools and their impact on patient adherence are necessary before these tools are professionally recommended to patients.

Acknowledgments

The authors are grateful to Roche-Chugai, as this work was initiated during continued medical education sessions on how to conduct scientific study organized by Roche-Chugai. Roche-Chugai had no role in the study design or the collection, analysis, or interpretation of the data; writing of the manuscript; or the decision to submit the manuscript for publication. No specific funding was received from anybody in the public, commercial, or non-profit sectors to carry out this study. The logistics of the investigators’ meetings were financially supported by Roche-Chugai, which had no role in the design, execution, analysis, or interpretation of the study.

Authors' Contributions

MM, EB, TB, CR, BC, MP, CL, PVS, AT, MET, and ARW were involved in conceptualizing or designing the study. MM, BC, CR, MP, and EB were involved in data collection. MM, EB, ADW, MET, and TB were involved in data analysis and interpretation. MM, ARW, and MET were responsible for drafting the manuscript. MM, EB, TB, CR, BC, MP, CL, PVS, AT, MET, and ARW provided critical revisions to the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection methods involving an anonymous self-questionnaire for patients and an anonymous and independent medical questionnaire for the rheumatologist in charge.
[DOCX File , 21 KB - jmir_v23i1e19998_app1.docx]

References

1. Guillemin F, Saraux A, Guggenbuhl P, Roux CH, Fardellone P, Le Bihan E, et al. Prevalence of rheumatoid arthritis in France: 2001. Ann Rheum Dis 2005 Oct;64(10):1427-1430 [FREE Full text] [doi: 10.1136/ard.2004.029199] [Medline: 15800010]
2. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
3. Liu W, Huang C, Wang C, Lee K, Lin S, Kuo H. A mobile telephone-based interactive self-care system improves asthma control. Eur Respir J 2011 Feb;37(2):310-317 [FREE Full text] [doi: 10.1183/09031936.00000810] [Medline: 20562122]
4. Charpentier G, Benhamou P, Dardari D, Clergeot A, Franc S, Schaepelynck-Belicar P, TeleDiab Study Group. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). Diabetes Care 2011 Mar 25;34(3):533-539 [FREE Full text] [doi: 10.2337/dc10-1259] [Medline: 21266648]
5. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 01;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
6. Gossec L, Cantagrel A, Soubrier M, Berthelot J, Joubert J, Combe B, et al. An e-health interactive self-assessment website (Sanoia) in rheumatoid arthritis. A 12-month randomized controlled trial in 320 patients. Joint Bone Spine 2018 Dec;85(6):709-714. [doi: 10.1016/j.jbspin.2017.11.015] [Medline: 29246532]
7. Salaffi F, Carotti M, Ciapetti A, Di Carlo M, Gasparini S, Farah S, et al. Effectiveness of a telemonitoring intensive strategy in early rheumatoid arthritis: comparison with the conventional management approach. BMC Musculoskelet Disord 2016 Apr 02;17:146. [doi: 10.1186/s12891-016-1002-2] [Medline: 27038788]
8. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
9. El Miedany Y, El Gaafary M, Youssef S, Bahlas S, Almedany S, Ahmed I, et al. Toward electronic health recording: evaluation of electronic patient-reported outcome measures system for remote monitoring of early rheumatoid arthritis. J Rheumatol 2016 Dec;43(12):2106-2112. [doi: 10.3899/jrheum.151421] [Medline: 27633823]
10. Aletaha D, Neogi T, Silman A, Funovits J, Felson D, Bingham C, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010 Sep;62(9):2569-2581 [FREE Full text] [doi: 10.1002/art.27584] [Medline: 20872595]
11. Baromètre du numérique 2018. CRÉDOC. 2018. URL: https://www.credoc.fr/publications/barometre-du-numerique-2018 [accessed 2021-01-14]
12. Eysenbach G. The impact of the internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [doi: 10.3322/canjclin.53.6.356] [Medline: 15224975]
13. Hansen AH, Broz J, Claudi T, Årsand E. Relations between the use of electronic health and the use of general practitioner and somatic specialist visits in patients with type 1 diabetes: cross-sectional study. J Med Internet Res 2018 Nov 07;20(11):e11322 [FREE Full text] [doi: 10.2196/11322] [Medline: 30404766]
14. Zhang Y, Lauche R, Sibbritt D, Olaniran B, Cook R, Adams J. Comparison of health information technology use between American adults with and without chronic health conditions: findings from the National Health Interview Survey 2012. J Med Internet Res 2017 Oct 05;19(10):e335 [FREE Full text] [doi: 10.2196/jmir.6989] [Medline: 28982644]
15. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin Shabr F, et al. Association of online health information-seeking behavior and self-care activities among type 2 diabetic patients in Saudi Arabia. J Med Internet Res 2015 Aug 12;17(8):e196 [FREE Full text] [doi: 10.2196/jmir.4312] [Medline: 26268425]

16. Mouelhi Y, Alessandrini M, Pauly V, Dussol B, Gentile S. Internet and social network users' profiles in renal transplant recipients in France. BMC Nephrol 2017 Aug 03;18(1):259 [FREE Full text] [doi: 10.1186/s12882-017-0670-y] [Medline: 28768480]
17. Chrischilles E, Hourcade J, Doucette W, Eichmann D, Gryzlak B, Lorentzen R, et al. Personal health records: a randomized trial of effects on elder medication safety. J Am Med Inform Assoc 2014;21(4):679-686 [FREE Full text] [doi: 10.1136/amiajnl-2013-002284] [Medline: 24326536]
18. Mollard E, Michaud K. A mobile app with optical imaging for the self-management of hand rheumatoid arthritis: pilot study. JMIR Mhealth Uhealth 2018 Oct 29;6(10):e12221 [FREE Full text] [doi: 10.2196/12221] [Medline: 30373732]
19. Vogt F, Seidl F, Santarpino G, van Griensven M, Emmert M, Edenharter G, et al. Healthcare IT utilization and penetration among physicians: novel IT solutions in healthcare - use and acceptance in hospitals. Eur Surg Res 2018 Jul 26;59(1-2):100-113. [doi: 10.1159/000490241] [Medline: 30048992]
20. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German mobile apps in rheumatology: review and analysis using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
21. Blenner SR, Köllmer M, Rouse AJ, Daneshvar N, Williams C, Andrews LB. Privacy policies of Android diabetes apps and sharing of health information. JAMA 2016 Mar 08;315(10):1051-1052. [doi: 10.1001/jama.2015.19426] [Medline: 26954415]
22. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]

Abbreviations

DMARDs: disease-modifying antirheumatic drugs
OR: odds ratio
RA: rheumatoid arthritis

Edited by G Eysenbach; submitted 08.05.20; peer-reviewed by S Sankaran, D Pförringer, C Fernández, V Strotbaum; comments to author 12.06.20; revised version received 11.07.20; accepted 11.11.20; published 29.01.21.

Please cite as:
Magnol M, Eleonore B, Claire R, Castagne B, Pugibet M, Lukas C, Tournadre A, Vergne-Salle P, Barnetche T, Truchetet ME, Ruyssen-Witrand A
Use of eHealth by Patients With Rheumatoid Arthritis: Observational, Cross-sectional, Multicenter Study
J Med Internet Res 2021;23(1):e19998
URL: http://www.jmir.org/2021/1/e19998/
doi:10.2196/19998
PMID:33512320

Original Paper

Content-Based Recommender Support System for Counselors in a Suicide Prevention Chat Helpline: Design and Evaluation Study

Salim Salmi[1], MSc; Saskia Mérelle[2], PhD; Renske Gilissen[2], PhD; Willem-Paul Brinkman[3], PhD

[1]Department of Stochastics, Centrum Wiskunde & Informatica, Amsterdam, Netherlands

[2]113 Suicide Prevention, Amsterdam, Netherlands

[3]Interactive Intelligence, Delft University of Technology, Delft, Netherlands

Corresponding Author:
Salim Salmi, MSc
Department of Stochastics
Centrum Wiskunde & Informatica
Science Park 123
Amsterdam, 1098 XG
Netherlands
Phone: 31 640673474
Email: s.salmi@cwi.nl

Abstract

Background: The working environment of a suicide prevention helpline requires high emotional and cognitive awareness from chat counselors. A shared opinion among counselors is that as a chat conversation becomes more difficult, it takes more effort and a longer amount of time to compose a response, which, in turn, can lead to writer's block.

Objective: This study evaluates and then designs supportive technology to determine if a support system that provides inspiration can help counselors resolve writer's block when they encounter difficult situations in chats with help-seekers.

Methods: A content-based recommender system with sentence embedding was used to search a chat corpus for similar chat situations. The system showed a counselor the most similar parts of former chat conversations so that the counselor would be able to use approaches previously taken by their colleagues as inspiration. In a within-subject experiment, counselors' chat replies when confronted with a difficult situation were analyzed to determine if experts could see a noticeable difference in chat replies that were obtained in 3 conditions: (1) with the help of the support system, (2) with written advice from a senior counselor, or (3) when receiving no help. In addition, the system's utility and usability were measured, and the validity of the algorithm was examined.

Results: A total of 24 counselors used a prototype of the support system; the results showed that, by reading chat replies, experts were able to significantly predict if counselors had received help from the support system or from a senior counselor (P=.004). Counselors scored the information they received from a senior counselor (M=1.46, SD 1.91) as significantly more helpful than the information received from the support system or when no help was given at all (M=–0.21, SD 2.26). Finally, compared with randomly selected former chat conversations, counselors rated the ones identified by the content-based recommendation system as significantly more similar to their current chats (β=.30, P<.001).

Conclusions: Support given to counselors influenced how they responded in difficult conversations. However, the higher utility scores given for the advice from senior counselors seem to indicate that specific actionable instructions are preferred. We expect that these findings will be beneficial for developing a system that can use similar chat situations to generate advice in a descriptive style, hence helping counselors through writer's block.

(J Med Internet Res 2021;23(1):e21690) doi:10.2196/21690

KEYWORDS

suicide prevention; content based recommender system; chat corpus; crisis line; sentence embedding; suicide; mental health

Introduction

Worldwide, helplines have been set up to help individuals who are struggling with suicidal thoughts. These helplines are a preventive service to reduce the suicidal ideation or behavior of help-seekers [1]. These help-seekers can contact trained volunteers and professionals (counselors) who can listen to them and assist them with their problems relating to suicide.

Historically, people have been able to contact these helplines over the telephone, but with the advent of the internet, chat services have become increasingly popular. Compared with telephone helplines, online chat helplines show approximately the same beneficial effects [2]. Help-seekers mention several reasons for using counseling through an online chat rather than a traditional phone call, such as privacy and the slow deliberate nature of online chatting [3-6]. In the Netherlands, the 113 Suicide Prevention service saw the number of conversations increase to more than 35,000 via telephone and more than 57,000 via online chat in 2018, an increase of 33% from 2017. However, this increase resulted in a higher need for counselors as well. Because of the difficult nature of crisis counseling, suicide prevention helplines often have difficulty retaining counselors [7].

Studies have indicated that technology can support chat line operators in executing cognitive tasks. For example, in the related field of commercial telephone and chat customer support, there are various supportive technologies developed for operators [8-10]. However, in computing research aimed at suicide prevention, most work focuses on the prediction and detection of suicidal behavior [11,12], while only a few studies have examined assisting online counselors; this could be beneficial, though. Salmi [13] has identified several difficulties that counselors encounter in their work. First, the counselor has to take in a large amount of information about the help-seeker. Here, counselors could be supported in understanding a help-seeker's history without having to read large portions of transcripts. Dinakar et al [14], therefore, have created a support system prototype for text-based crisis counseling called Fathom. Fathom uses visualizations based on topic modeling to provide information at a glance. In comparison to a control interface without a visualization aspect, Fathom was preferred by counselors when eliciting a list of issues and a conversation summary. Another difficulty is that the counselor must be aware of the conversation quality. In this respect, Althoff et al [15] compared the chat conversations of more and less successful counselors with natural language processing techniques to discover the quality differences, defining actionable strategies to improve conversation quality. For example, they showed that more successful counselors spend a longer time exploring solutions, while less successful counselors spend more time defining the problems.

Finally, the complexity and severity of help-seekers' situations may lead to writer's block in counselors. Although not directly related to the suicide prevention domain, Isbister et al [16] have designed a helper agent for human-human interaction. When a conversation lags, the agent suggests topics for the conversation pair to talk about and, thereby, the agent is generally able to make positive contributions to the chat.

In situations where counselors experience writer's block, a straightforward solution would be to approach a senior colleague for help. These senior counselors can read along and describe in as much detail as necessary how they would respond to the help-seeker. However, this requires availability and time from a colleague, and this is not always possible. Responding quickly is important in life-threatening situations, and counselors cannot always wait for somebody to become available. We also suspect that an approach such as suggesting topics to keep a conversation going [16] or providing a conversation summary [12] would not be optimal in difficult situations where counselors have to de-escalate a suicide-related crisis. This paper, therefore, presents a system that uses natural language processing techniques to provide support for counselors in difficult chat conversations. The system recommends parts of similar, previous chat situations for the counselor to draw inspiration from, which might be able to reduce their writer's block. This paper also evaluates the designed support system by comparing it with 1) written, general advice from a senior counselor and 2) receiving no additional help during chats. The system's usability and utility, along with the validity of the algorithm used, were also examined.

Methods

Design

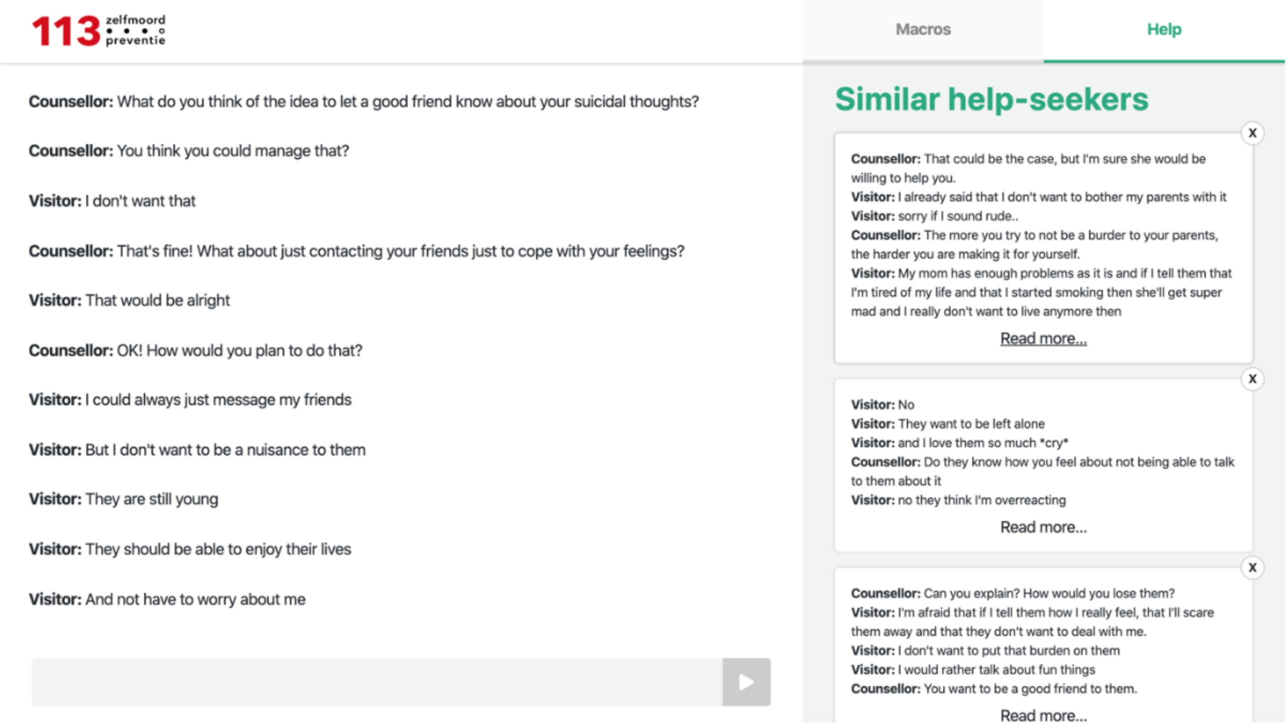
We used a within-subject design to evaluate the impact and usefulness of similar chat situations that could be used as inspiration. In the study, the counselor wrote a chat reply to a simulated chat that was interrupted as a difficult situation. The counselor took part in 3 simulations: 1) the counselor received parts of similar chats from a support system, 2) the counselor received written advice from an experienced counselor, and 3) the counselor received no additional help. A questionnaire was used to measure the support system's usability. Finally, we evaluated the validity of the similarity of the generated chats by testing the algorithm in a small additional experiment with a within-subject design.

The current study received ethical approval from the TU Delft University research ethics committee (id: 688). Before starting the data collection, the experimental setup was also preregistered on the Open Science Framework [17].

Recommender Support System

For the study, we developed a system recommending the transcripts of similar previous chat conversations to a counselor based on the content of the counselor's current chat conversation. Figure 1 shows a chat window on the left and the support system interface on the right. The support system shows the top 10 most similar chat messages, which the counselor could click to read them in their entirety.

Figure 1. Interface support system (right). Content translated from Dutch.



Counsellor: What do you think of the idea to let a good friend know about your suicidal thoughts?

Counsellor: You think you could manage that?

Visitor: I don't want that

Counsellor: That's fine! What about just contacting your friends just to cope with your feelings?

Visitor: That would be alright

Counsellor: OK! How would you plan to do that?

Visitor: I could always just message my friends

Visitor: But I don't want to be a nuisance to them

Visitor: They are still young

Visitor: They should be able to enjoy their lives

Visitor: And not have to worry about me

Counsellor: That could be the case, but I'm sure she would be willing to help you.
Visitor: I already said that I don't want to bother my parents with it
Visitor: sorry if I sound rude..
Counsellor: The more you try to not be a burder to your parents, the harder you are making it for yourself.
Visitor: My mom has enough problems as it is and if I tell them that I'm tired of my life and that I started smoking then she'll get super mad and I really don't want to live anymore then

Visitor: No
Visitor: They want to be left alone
Visitor: and I love them so much *cry*
Counsellor: Do they know how you feel about not being able to talk to them about it
Visitor: no they think I'm overreacting

Counsellor: Can you explain? How would you lose them?
Visitor: I'm afraid that if I tell them how I really feel, that I'll scare them away and that they don't want to deal with me.
Visitor: I don't want to put that burden on them
Visitor: I would rather talk about fun things
Counsellor: You want to be a good friend to them.

A corpus of chat conversations between help-seekers and counselors was used to find similar previous chat situations. We used the corpus from 113 Suicide Prevention in the Netherlands. This corpus contained 7 months of chats spanning from March 2018 to September 2018. The chat data were first filtered, removing all chats that had less than 20 interactions. In total, we used 17,773 chats. Furthermore, any special symbols in the messages were cleaned, and capital letters were replaced by lowercase letters.

Because the chats each contained multiple problems, we used a sliding window algorithm to scan for relevant chat segments instead of comparing complete chats. This algorithm created sets of chat messages, starting with the first 5 messages. The next set removed the first message in the window and added the sixth message; this process was repeated to create every possible set of 5 subsequent messages in a chat. The sliding window algorithm was then used to create the chat segments for the entire corpus.

We used an embedding algorithm to compute the similarity. For each chat segment, an embedding was created using smooth inverse frequency [18], which takes a weighted average of the word embeddings for each word in the text of the window corresponding to the inverse of the frequency of the word in the corpus. This resulted in less meaningful words receiving a lower weight. To create word embeddings, Mikolov et al [19] developed an algorithm dubbed Word2Vec, improving previous methods [20]. The word embeddings we used were obtained from the COOSTO Word2Vec model [21], a model developed using Dutch social media and blog posts. A window of 5 messages resulted in 1,286,659 embeddings, which were stored alongside the corresponding chat and window positions.

When a counselor in an ongoing conversation requested similar chat conversations, a single smooth inverse frequency embedding was created using the same steps as with the corpus, except only the last 5 messages of the ongoing conversation were used. This embedding was then compared with the corpus embeddings through a cosine similarity. Ten windows with the highest similarity were recommended to the counselor.

Difficult Chats

We used 6 chats for the experiment to cover several difficult situations: a situation where a help-seeker 1) was in a dangerous location and had withheld this from the counselor; 2) did not want to inform anybody in their environment of their suicidality because they felt like it would put a burden on others; 3) was afraid of people in their environment not understanding their problems; 4) tried to look for help but was not believed; 5) was excessively rude; and 6) had to contact a psychologist.

Participants

Counselor and expert recruitment, as well as conducting the experiment, happened at 113 Suicide Prevention. In total, 24 counselors participated. On average, the participants' age was 27 years old, and 79% were female. Only counselors who were interns, volunteers, or trainees were eligible to participate. Each counselor met all the components and conditions of the evaluation.

Measures

The perceived utility was assessed with the following question: "How, in your opinion, did the extra information help you with coming up with your response?". The counselors graded each support type on a fixed interval scale from –3 to 3, where –3 indicated the extra information was hindering, 0 indicated the information was neutral, and 3 indicated the information was useful.

To measure usability, the counselors were asked to fill out the System Usability Scale questionnaire [22]; this is a validated

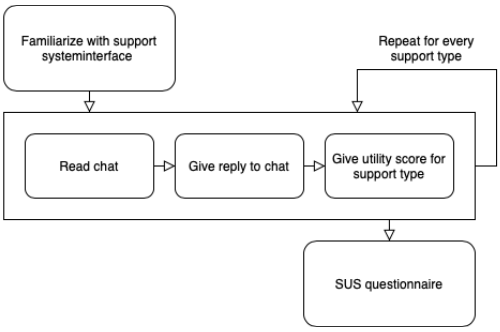
10-item questionnaire with a 5-point scale ranging from "Strongly disagree" to "Strongly agree."

To measure the validity of the algorithm, the counselors used a 7-point fixed interval scale to indicate how much they agreed with the following statement: "The problem in the matched chat is the same as the problem in the ongoing chat." A score of 1 meant the counselor did not agree, whereas a score of 7 meant they did agree.

Procedure

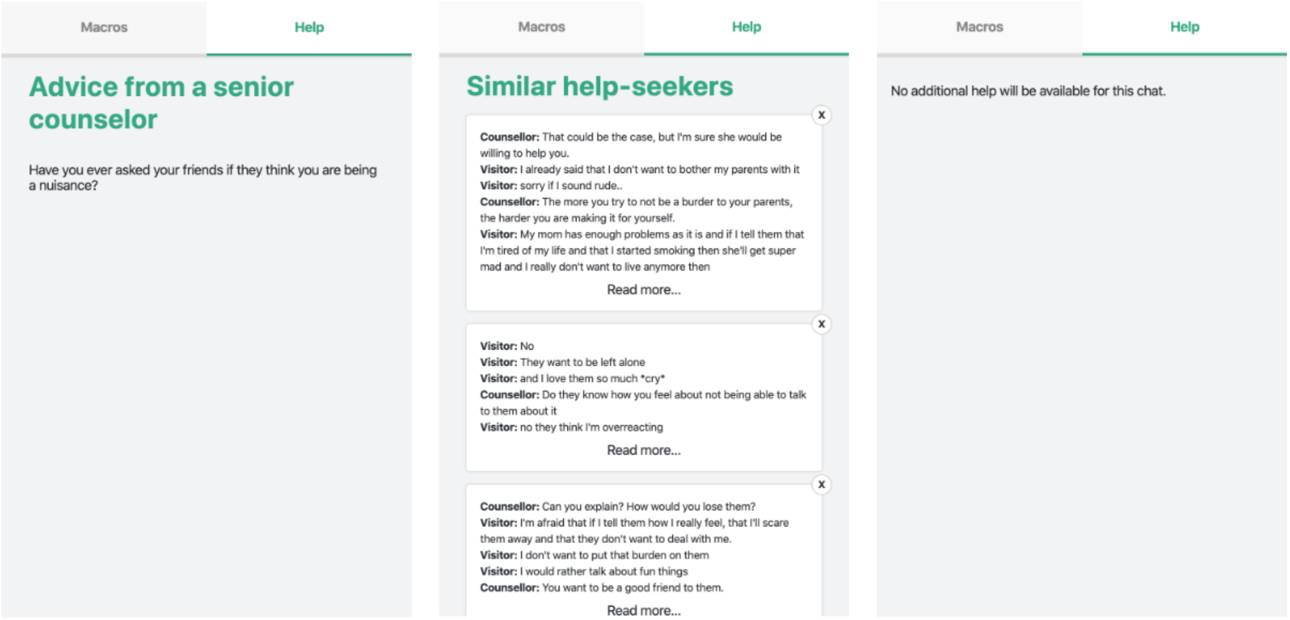
The counselors used a test environment with simulated chats. The experiment consisted of 2 parts. Figure 2 shows a diagram of the procedure for the first part. Before the experiment, the counselor had 5 minutes to explore and familiarize themself with the support system.

Figure 2. Procedure diagram of the first part of the experiment. SUS: System Usability Scale.



Part 1 consisted of a simulated environment where the counselor read and reacted to 3 simulated chats, one after the other. The support information was contained in an extra tab called "Help." Figure 3 shows each support type. Each counselor had the same amount of time to read the chat. To simulate a real situation, each counselor had a 2-minute window to reply. The counselor could not access the support tab before the 2-minute timer started. Directly after the counselor submitted their reply to a chat, they were asked to rate the utility of the support type. These steps were repeated for each condition. Therefore, the participants reacted to 3 chats in total. The chats, support types, and combinations were counterbalanced for the 24 participants. This part ended with the System Usability Scale questionnaire being used to measure the usability of the support system.

Figure 3. Conditions of the experiment: senior counselor written advice (left); support system (center); no additional help (right). Content translated from Dutch.



Advice from a senior counselor

Have you ever asked your friends if they think you are being a nuisance?

Similar help-seekers

Counsellor: That could be the case, but I'm sure she would be willing to help you.
Visitor: I already said that I don't want to bother my parents with it
Visitor: sorry if I sound rude..
Counsellor: The more you try to not be a burder to your parents, the harder you are making it for yourself.
Visitor: My mom has enough problems as it is and if I tell them that I'm tired of my life and that I started smoking then she'll get super mad and I really don't want to live anymore then

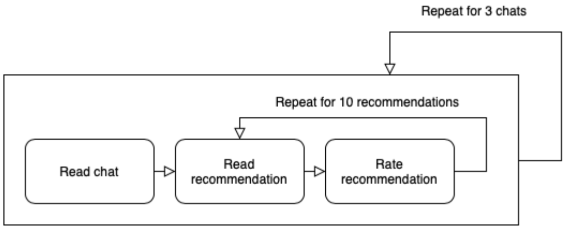
Visitor: No
Visitor: They want to be left alone
Visitor: and I love them so much *cry*
Counsellor: Do they know how you feel about not being able to talk to them about it
Visitor: no they think I'm overreacting

Counsellor: Can you explain? How would you lose them?
Visitor: I'm afraid that if I tell them how I really feel, that I'll scare them away and that they don't want to deal with me.
Visitor: I don't want to put that burden on them
Visitor: I would rather talk about fun things
Counsellor: You want to be a good friend to them.

No additional help will be available for this chat.

Part 2 recorded the measurements for evaluating the validity of the algorithm. Figure 4 shows a diagram of the procedure. The left side of the screen contained the transcript of an ongoing chat. The right side of the screen showed 10 chat segments. Half of these segments were randomly selected, and the other half was matched to the ongoing chat using the embedding algorithm. Below each of the segments was a fixed interval scale from 1 to 7 where the counselor rated the degree to which that chat segment related to the ongoing chat. To enhance generalization, the participants did this for the transcripts of 3 different ongoing chats. Therefore, in total, a participant rated 30 segments.

Figure 4. Procedure diagram of the second part of the experiment.



Data Preparation

Eight experts labeled the reply of the counselor with the type of help (condition) that the expert assumed that the counselor had received. To prevent expert bias, each expert judged all the counselor responses. Furthermore, a reliability analysis for the items of the System Usability Scale questionnaire showed an acceptable level of consistency, with a Cronbach alpha of .89. Therefore, the System Usability Scale items were compiled into a single score.

Analysis

The noticeable difference in counselor outputs was analyzed using generalized mixed-effects analyses [23] to predict the outcome variable support type based on the label the expert assigned to the counselor reply. The analyses were done by comparing 2 support type conditions at a time, thereby excluding the data from 1 of the 3 support type conditions. The models fitted on the remaining 2 conditions hence assumed a binomial distribution. Each model was compared with a null model that did not include an expert label as a fixed effect. Because the test was conducted 3 times, a Bonferroni correction [24] was used to set the significance threshold at .016. In addition, crossed random effects were used with random intercepts for the counselor and expert. Furthermore, for each support type, the utility ratings were analyzed using a one-sample *t* test to examine whether the rating deviated from the neutral zero score on the scale.

To examine the validity of the algorithm, a linear mixed-effects analysis was performed on the counselor’s rating of the similarity between the chat segment and the ongoing chat. As a two-level fixed effect the analysis included the recommendation method, that is, randomly selected versus selected by embedding algorithm. Furthermore, the ongoing chat was added as a three-level fixed variable because the quality of the suggestions was assumed to depend on the specific chat. As a random effect, the intercepts for counselors were used.

Anonymized data and R scripts are available online [25].

Results

Noticeable Difference in Counselor Output

Table 1 shows the effect of support type on the outcome measure of the expert label. The first row shows that the expert label significantly predicts the support type, when the data of no support condition was left out. In other words, the experts could tell the difference between replies given with the support system and replies with help from a senior counselor. Table 2 shows that when the expert labeled the counselor’s response as having received help from the support system, the counselor was 0.47 times less likely to have received the senior counselor support. This effect is further illustrated when looking at the confusion matrix of these conditions, as shown in Figure 5. However, no significant difference was found between the no support condition and any of the other conditions.

Table 1. Results of the comparison between null model and full models that included the expert label as a fixed effect to predict support type counselors had received when writing their reply (n=356).

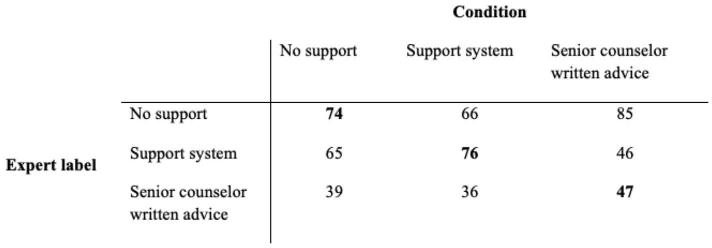
Outcomes data included in analysis	$\chi 2$ (*df*)	*P* value
Support system and senior counselor written advice	11.31 (2)	.004
No support and support system	1.44 (2)	.49
No support and senior counselor written advice	4.78 (2)	.09

Table 2. Fixed effect of the expert label for the model of support system and senior counselor written advice.

Parameter	OR[a]	Standard error	z Value	*P* value
Intercept	1.29	0.16	1.54	.12
Support system	0.47	0.25	–3.04	.002
Senior counselor written advice	1.01	0.28	0.049	.96

[a] OR: odds ratio.

Figure 5. Confusion matrix for expert labeling of counselor responses.



		Condition		
		No support	Support system	Senior counselor written advice
Expert label	No support	**74**	66	85
	Support system	65	**76**	46
	Senior counselor written advice	39	36	**47**

Utility

The results of the utility ratings are shown in Table 3. The mean score of the support system was –0.21 (SD 2.26) and did not significantly deviate from 0, indicating that there was neither a hindering nor a helping effect experienced by the counselors. However, the mean utility score of the written advice from a senior counselor was 1.46 (SD 1.91) and significantly deviated from 0. This suggests that the written advice was perceived as helpful. It is noteworthy that the support system had a high variance, suggesting that the counselor's opinion on the utility was divided.

Table 3. One-sample t test for counselor utility ratings per support types (n=24).

Support type	Mean (SD)	95% CI	t_{df}	*P* value
Support system	–0.21 (2.26)	–0.84 to 0.43	-0.68_{23}	.5
Counselor written advice	1.46 (1.91)	0.87 to 2.04	5.17_{23}	<.001
No support	–0.21 (0.95)	–0.62 to 0.2	-1.04_{23}	.31

Usability

The mean score of the support system for the System Usability Scale questionnaire was 71, with a 95% confidence interval of 63-78. According to Bangor et al [26], this score can be classified as "good" based on an adjective rating scale.

Validity of the Algorithm

How the chat segments were selected (randomly vs by the embedding algorithm) significantly predicted the rating counselors gave on the chat segment's similarity to the ongoing chat, β=.30, t(7.66), *P*<.001. This means that counselors could tell the difference between the random chats and those generated by the support system. The suggestions from the algorithm increased the similarity rating given by counselors from an average of 2.35 to an average of 3.42 (difference of 1.07).

Discussion and Conclusions

In the current study, we evaluated a prototype support system to assist chat counselors in suicide prevention helplines by providing inspiration from previous chats. The results show that counselors gave different answers depending on whether they received help from the support system or from a senior colleague. Upon inspection, the replies given by the counselors who received written advice from a senior colleague were, for the most part, copied directly and with little to no alterations made. Replies from counselors using the support system were more varied. This could be a possible explanation for the noticeable difference. However, we could not find a significant result for the no-help condition, which also had varied replies. Additionally, we observed that written advice from a senior counselor was given a significantly higher utility score than the other conditions; this suggests that the counselors value short actionable information that is highly accurate to the situation and that is given by someone with expertise. Gunaratne et al [27] have observed similar findings in their study on the effects of expert advice and social comparison on decision making for retirement savings; they showed that expert advice helped people make better decisions, whereas social comparison was seen as a useful mechanism to keep people from deviating too far from the mean and, hence, make safe decisions. However, both of these conditions outperformed a control condition where no additional information was provided.

The main contribution of our study is the idea of retrieving inspiration from a conversation corpus. Other support systems for chats [28-30] have used topics to assist the conversation. Compared with these methods, our approach for combating writer's block in a counseling conversation is novel. Furthermore, an experimental design was used to compare this supportive technology with advice from a senior colleague, showing how the two differ.

Some limitations should be considered regarding the findings and their implications. We used chat transcripts of conversations with situations that previous counselors found difficult to evaluate. However, this might not cause writer's block for every participant because not every counselor will have problems with the same situations. For writer's block to occur naturally, the system would have to be tested in live chats. This was, however, not possible because of the ethical constraints of deploying an unevaluated system in a possible crisis situation. Furthermore, the specification, development, and evaluation were done in the context of counselors working at 113 Suicide Prevention in the Netherlands, with a limited number of counselors. The support system should also be tested in different helplines and with a larger sample size.

We have outlined two major directions for future work. First, the recommendation mechanism could be improved in different ways. This study, as well as other related works such as recommenders for creativity [31] and scientific writing [32], relies on topic modeling and bag-of-word models to find recommendations. Encoding text using attention-based models [33], such as BERT [34], have been shown to perform well on various natural language processing tasks, including semantic sentence similarity for conversation data [28]. These methods could be applied to improve the recommendations to find more relevant and similar examples, which we expect will increase the perceived utility. Additionally, curating the corpus can help denoise the dataset and improve the recommendations. This can also give counselors the knowledge that the information comes from a subset of quality chats, thereby acting on the persuasive principle of authority as outlined by Cialdini [29]. Lastly, there is also an opportunity to apply active learning methods by adding positive labels to the recommendations that the counselors interacted with or explicitly marked as useful [30].

Second, the findings show that the embedding algorithm found similar chats and that written advice from senior counselors had high utility. Compared to the Gunaratne et al study [27], the main difference to the setup of our study is that the social comparison condition provided information as an average; this indicates that refining the output of the support system recommendations to be more instructional could be a possible direction for improving the system. To combine both the extensive coverage of a chat corpus and the high utility of curated written advice, clustering could be used, that is, grouping similar chats together based on a similarity metric and curating the labels based on these clusters. Derrar [35] uses clustering to automate the annotation of customer service chat messages. A similar approach could be used to annotate the chat corpus to create a taxonomy of situations and advice, which then could emulate receiving written advice from a senior colleague. In other words, working together with experts, a set of advice could be formulated in advance for each specific situation. Next, a data-driven algorithm could be trained to classify chats according to categories of the taxonomy, consequently providing counselors with expert advice associated with the category and making the expert advice situation relevant. This approach would be most suitable for assisting counselors with frequently occurring tasks, as these would be the most likely cases to be included in the taxonomy. The focus of the support system might therefore shift from an inspiration source to a system that could reduce workload. Alternatively, the field of conversational information retrieval has explored multiple methods that could be applied to the task presented in this paper. For example, Qiu et al [36] combined both information retrieval methods and generation based models to create a chat bot trained using existing customer service chat logs. These techniques could potentially also be used to allow the system to generate proposal responses that counselors could consider using in their chats with help-seekers.

In conclusion, the current study shows a possible method to provide inspiration during chat counseling in a helpline for suicide prevention and how this supportive technology compares with human assistance. A support system may be a relief for counselors as they handle many cognitively difficult situations. In addition, supportive technology seems useful for helplines to better deal with busy periods, to provide a safety-net for junior counselors, and to help sustain counselors.

Acknowledgments

The authors would like to thank the counselors of 113 Suicide Prevention who participated in the experiment. The authors would also like to thank the floor managers of 113 Suicide Prevention who provided the senior counselor comments and helped select the difficult chat situations.

Conflicts of Interest

None declared.

References

1. Gould MS, Kalafat J, Harrismunfakh JL, Kleinman M. An evaluation of crisis hotline outcomes. Part 2: Suicidal callers. Suicide and Life-Threatening Behavior 2007 Jun;37(3):338-352. [doi: 10.1521/suli.2007.37.3.338] [Medline: 17579545]
2. Mokkenstorm JK, Eikelenboom M, Huisman A, Wiebenga J, Gilissen R, Kerkhof AJFM, et al. Evaluation of the 113Online Suicide Prevention Crisis Chat Service: Outcomes, Helper Behaviors and Comparison to Telephone Hotlines. Suicide and Life-Threatening Behavior 2016 Aug 19;47(3):282-296. [doi: 10.1111/sltb.12286]
3. Evans WP, Davidson L, Sicafuse L. Someone to Listen: Increasing Youth Help-Seeking Behavior Through a Text-Based Crisis Line for Youth. J. Community Psychol 2013 Mar 21;41(4):471-487. [doi: 10.1002/jcop.21551]
4. Suler J. The psychology of text relationships. Online Counseling (Second Edition) 2011:21-53. [doi: 10.1016/B978-0-12-378596-1.00002-2]
5. Dowling M, Rickwood D. Online Counseling and Therapy for Mental Health Problems: A Systematic Review of Individual Synchronous Interventions Using Chat. Journal of Technology in Human Services 2013 Jan;31(1):1-21. [doi: 10.1080/15228835.2012.728508]
6. Fukkink R, Hermanns J. Counseling children at a helpline: chatting or calling? J. Community Psychol 2009 Nov;37(8):939-948. [doi: 10.1002/jcop.20340]
7. Pratt ME. The future of volunteers in crisis hotline work. 2013. URL: http://d-scholarship.pitt.edu/16489/ [accessed 2020-12-11]

8. Acorn TL, Walden SH. SMART: support management automated reasoning technology for compaq customer service. In: IAAI'92: Proceedings of the fourth conference on Innovative applications of artificial intelligence. 1992 Presented at: The Fourth Conference on Innovative applications of artificial intelligence; 1992; San Jose, California p. 3-18.
9. Goldberg RG, Rosinski RR. Automated natural language understanding customer service system. Google patents. 1999. URL: https://patents.google.com/patent/US5895466A/en [accessed 2020-12-11]
10. Ignacio Madrid R, Van Oostendorp H, Puerta Melguizo MC. The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order. Computers in Human Behavior 2009 Jan;25(1):66-75. [doi: 10.1016/j.chb.2008.06.005]
11. Walsh CG, Ribeiro JD, Franklin JC. Predicting Risk of Suicide Attempts Over Time Through Machine Learning. Clinical Psychological Science 2017 Apr 11;5(3):457-469. [doi: 10.1177/2167702617691560]
12. Coppersmith G, Leary R, Crutchley P, Fine A. Natural Language Processing of Social Media as Screening for Suicide Risk. Biomed Inform Insights 2018 Aug 27;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
13. Salmi S. Context-based recommender system to provide cognitive support to online chat counsellors in the Helpline of 113 Suicide Prevention. 2019 Nov 26. URL: https://repository.tudelft.nl/islandora/object/uuid:cd6e857c-7241-48be-b71b-59371401a640 [accessed 2020-12-11]
14. Dinakar K, Chen J, Lieberman H, Picard R, Filbin R. Mixed-Initiative Real-Time Topic Modeling & Visualization for Crisis Counseling. In: IUI '15: Proceedings of the 20th International Conference on Intelligent User Interfaces. 2015 Presented at: IUI'15 20th International Conference on Intelligent User Interfaces; 2015; Atlanta Georgia USA. [doi: 10.1145/2678025.2701395]
15. Althoff T, Clark K, Leskovec J. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. TACL 2016 Dec;4:463-476. [doi: 10.1162/tacl_a_00111]
16. Isbister K, Nakanishi H, Ishida T, Nass C. Helper agent: designing an assistant for human-human interaction in a virtual meeting space. In: CHI '00: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 2000 Apr Presented at: CHI00: Human Factors in Computing Systems; 2000; The Hague, The Netherlands. [doi: 10.1145/332040.332407]
17. Salmi S, Brinkman WP. Evaluation inspiration support system for suicide crisis counseling. Open Science Framework. 2019 May 13. URL: https://osf.io/9gu2y [accessed 2019-05-13]
18. Arora S, Liang Y, Ma T. A simple but tough-to-beat baseline for sentence embeddings. In: 5th International Conference on Learning Representations. 2017 Presented at: 5th International Conference on Learning Representations; 2017; Toulon, France.
19. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems 26 (NIPS 2013). 2013 Presented at: Advances in neural information processing systems 26; 2013; Lake Tahoe, Nevada, United States p. 3111-3119.
20. Baroni M, Dinu G, Kruszewski G. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014 Presented at: 2014 Association for Computational Linguistics; 2014; Baltimore, Maryland, USA p. 238-247. [doi: 10.3115/v1/p14-1023]
21. Nieuwenhuijse A. COOSTO Dutch Word2Vec Model. GitHub. URL: https://github.com/coosto/dutch-word-embeddings [accessed 2019-03-20]
22. Brooke J. SUS-A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability evaluation in industry. London: Taylor & Francis; 1996:189-194.
23. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Soft 2015 Oct 07;67(1):1-48. [doi: 10.18637/jss.v067.i01]
24. Abdi H. The Bonferonni and Šidák Corrections for Multiple Comparisons. In: Salkind NJ, editor. Encyclopedia of Measurement and Statistics volume 3. Newbury Park, California, USA: SAGE Publishing; 2007:103-107.
25. Salmi S, Mérelle S, Gilissen R, Renske B, Brinkman WP. Data and analysis for the publication: Content-Based Recommender Support System for Counselors in a Suicide Prevention Chat Helpline: Design and Evaluation Study.: 4TU.ResearchData; 2020. URL: https://doi.org/10.4121/13366739
26. Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of usability studies 2009 May;4(3):114-123.
27. Gunaratne J, Nov O. Influencing Retirement Saving Behavior with Expert Advice and Social Comparison as Persuasive Techniques. In: International Conference on Persuasive Technology. 2015 Presented at: International Conference on Persuasive Technology; 2015; Chicago, Illinois, USA p. 205-216. [doi: 10.1007/978-3-319-20306-5_19]
28. Yang Y, Yuan S, Cer D, Kong SY, Constant N, Pilar P, et al. Learning Semantic Textual Similarity from Conversations. In: Proceedings of the 3rd Workshop on Representation Learning for NLP. 2018 Presented at: The 3rd Workshop on Representation Learning for NLP; July 20, 2018; Melbourne, Australia p. 164-174. [doi: 10.18653/v1/w18-3022]
29. Cialdini RB. Authority. In: Influence: The psychology of persuasion. New York City, USA: Harper Collins Publishers Inc; 2007.

30. Cha Zhang, Tsuhan Chen. An active learning framework for content-based information retrieval. IEEE Trans. Multimedia 2002 Jun;4(2):260-268. [doi: 10.1109/tmm.2002.1017738]
31. Sielis GA, Mettouris C, Papadopoulos GA. A Context Aware Recommender System for Creativity Support Tools. Journal of Universal Computer Science 2011 Aug 1;17(12):1743-1763.
32. Achakulvisut T, Acuna DE, Ruangrong T, Kording K. Science Concierge: A Fast Content-Based Recommendation System for Scientific Publications. PLoS One 2016 Jul 6;11(7):e0158423 [FREE Full text] [doi: 10.1371/journal.pone.0158423] [Medline: 27383424]
33. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. 2017 Presented at: 31st Conference on Neural Information Processing Systems; December 2017; Long Beach, CA p. 6000-6010 URL: https://dl.acm.org/doi/pdf/10.5555/3295222.3295349 [doi: 10.5555/3295222.3295349]
34. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019 Presented at: Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2019; Minneapolis URL: https://arxiv.org/pdf/1810.04805.pdf [doi: 10.18653/v1/N19-1423]
35. Derrar HM. Clustering for the automatic annotation of customer service chat messages. 2019. URL: https://trepo.tuni.fi/bitstream/handle/123456789/27114/Derrar.pdf [accessed 2020-12-11]
36. Qiu M, Li FL, Wang S, Gao X, Chen Y, Zhao W, et al. AliMe Chat: A Sequence to Sequence and Rerank based Chatbot Engine. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017 Presented at: The 55th Annual Meeting of the Association for Computational Linguistics; 2017; Vancouver, Canada p. 498-503. [doi: 10.18653/v1/p17-2079]

Edited by G Eysenbach; submitted 23.06.20; peer-reviewed by A Foreman, J Han; comments to author 21.08.20; revised version received 04.09.20; accepted 28.10.20; published 07.01.21.

Please cite as:
Salmi S, Mérelle S, Gilissen R, Brinkman WP
Content-Based Recommender Support System for Counselors in a Suicide Prevention Chat Helpline: Design and Evaluation Study
J Med Internet Res 2021;23(1):e21690
URL: https://www.jmir.org/2021/1/e21690
doi:10.2196/21690
PMID:33410755

Original Paper

Gaining Insights Into the Estimation of the Circadian Rhythms of Social Activity in Older Adults From Their Telephone Call Activity With Statistical Learning: Observational Study

Timothée Aubourg[1,2,3], PhD; Jacques Demongeot[2,3,4], MD, PhD; Nicolas Vuillerme[2,3,4], PhD

[1]Orange Labs, Meylan, France

[2]Univ. Grenoble Alpes, AGEIS, Grenoble, France

[3]LabCom Telecom4Health, Univ. Grenoble Alpes & Orange Labs, Grenoble, France

[4]Institut Universitaire de France, Paris, France

Corresponding Author:
Timothée Aubourg, PhD
Orange Labs
Chemin du Vieux Chêne
Meylan
France
Phone: 33 0438429192
Email: timotheeaubourg@gmail.com

Abstract

Background: Understanding the social mechanisms of the circadian rhythms of activity represents a major issue in better managing the mechanisms of age-related diseases occurring over time in the elderly population. The automated analysis of call detail records (CDRs) provided by modern phone technologies can help meet such an objective. At this stage, however, whether and how the circadian rhythms of telephone call activity can be automatically and properly modeled in the elderly population remains to be established.

Objective: Our goal for this study is to address whether and how the circadian rhythms of social activity observed through telephone calls could be automatically modeled in older adults.

Methods: We analyzed a 12-month data set of outgoing telephone CDRs of 26 adults older than 65 years of age. We designed a statistical learning modeling approach adapted for exploratory analysis. First, Gaussian mixture models (GMMs) were calculated to automatically model each participant's circadian rhythm of telephone call activity. Second, k-means clustering was used for grouping participants into distinct groups depending on the characteristics of their personal GMMs.

Results: The results showed the existence of specific structures of telephone call activity in the daily social activity of older adults. At the individual level, GMMs allowed the identification of personal habits, such as morningness-eveningness for making calls. At the population level, k-means clustering allowed the structuring of these individual habits into specific morningness or eveningness clusters.

Conclusions: These findings support the potential of phone technologies and statistical learning approaches to automatically provide personalized and precise information on the social rhythms of telephone call activity of older individuals. Futures studies could integrate such digital insights with other sources of data to complete assessments of the circadian rhythms of activity in elderly populations.

(J Med Internet Res 2021;23(1):e22339) doi:10.2196/22339

KEYWORDS

circadian rhythms; phone call detail records; older population; statistics; machine learning

Introduction

Background

Circadian rhythms are endogenous processes that regulate the individual's activity over a 24-hour cycle. They are recognized to play a crucial role in nearly all biological and social aspects of the individual's life, as it is evidenced in his or her brain activity [1], rest activity [2], body temperature fluctuation [3], and rhythms of social interactions [4]. Such rhythms of activity, maintained over time, permit the individual to efficiently organize his or her physical and social activities at a daily scale. Such organization has played a significant role in the past from an evolutionary perspective for ensuring the survival of the human species [5]. Nowadays, circadian rhythms continue to be essential for our lives, giving us the ability to ensure a certain stability in the way we routinely interact with ourselves and with each other.

Furthermore, it is recognized that the circadian rhythms of activity significantly change with age (eg, see Hood and Amir, Steponenaite et al, Zhao et al, and Duffy et al [6-9] for recent reviews). At the endogenous level, the aging process may contribute to the occurrence of biological dysregulations that alter the internal clock of an older individual [10,11]. These changes may have potential adverse consequences on his or her body functioning (eg, see Hood and Amir [6] and Leng et al [12] for recent reviews), such as balance control alteration [13], sleep cycle disruptions [2,14], neurodegeneration [14], or cardiovascular complications [15]. Thus, with regard to our traditional health care system, it seems important for the health practitioner to assess the circadian rhythms of activity of his or her older patients in order to understand, and then to properly manage, the mechanisms of health issues that may occur in their life span.

Over the past decades, the assessment of the circadian rhythms of activity in humans has been addressed at the biological and physical levels through, respectively, the field of chronobiology (eg, see Cornelissen and Otsuka [16] and Otsuka et al [17] for two recent reviews) and that of actigraphy (eg, see Tazawa et al [18] and Schwab et al [19] for two recent reviews). On the whole, both of these approaches—chronobiology and actigraphy—have demonstrated the relevance of combining the use of wearable sensors with the implementation of statistical learning algorithms for automatically measuring the biological and gross motor activities of the individual at a daily scale. From a clinical perspective, these approaches may permit the detection in real time of the occurrence of daily risky situations that may affect the older individual's life, such as sedentary behavior [20], nocturnal activity [21], fall occurrences [22], or myocardial infarction [23]. Taken together, these approaches provide innovative and relevant ways to diagnose and manage the older adult's health status over time.

At this point, however, it is also important to mention that the daily rhythms of activity that are of a social nature are evidently not considered by such approaches as relying on biological and physical sources of data. At the clinical level, this issue is of significant importance given that biological circadian dysregulations can be associated with external social factors [24], such as inconsistent mealtime [25], shift work [26], or social isolation [27]. These social disruptions, maintained over time, may lead to a misalignment between biological and social rhythms, so-called "social jet lag" [28]. In the older adult, this misalignment can act as a retroactive feedback mechanism that contributes to the occurrence or worsening of particular symptoms or comorbidity factors, such as fatigue [2,29] and signs of anxiety and depression [30], that are involved in various age-related diseases [12,13,31].

Prior Work

In the field of health care monitoring for older adults, traditional approaches such as clinical questionnaires propose subjective, timely solutions for measuring and characterizing the circadian rhythms of social activity of the individual. Despite their widespread use in practice, this type of approach introduces limitations into their use for both clinical and research applications. One reason is related to the fact that the administration of questionnaires is time-consuming and requires the active participation of the patient for filling in a given questionnaire. These two constraints cannot be easily satisfied regarding specific vulnerable populations that are potentially not able to make such an effort because of their physical or psychological conditions. A second reason is because of the temporal scope of their results. Since these questionnaires are filled out at a set point in time, it is not possible to infer the general daily rhythms of the individual over a long period of time. A third reason is because of their subjective nature. Subjective responses that are provided for measuring objective temporal phenomena, such as the circadian rhythms of social activity, makes it difficult to ensure the precision of the questionnaires' results.

Modern technologies traditionally considered as everyday tools have the potential to get around these limits [32-34]. In particular, it has been recently demonstrated that the telephone may have the ability to passively, objectively, and continuously provide estimation of social rhythms through the analysis of its generated data over time [32,35-37]. Notably, two recent descriptive studies [32,38] have analyzed an 18-month data set of call detail records (CDRs) of 24 young adults. The results of these descriptive studies suggested that young participants seemed to hold their own personal circadian patterns for communicating with their social network. More importantly, these circadian patterns were found to persist over time. More recently, such a descriptive analysis was successfully reproduced, and got the same results, with an older population [39]. To the best of our knowledge, these preliminary results were the first ones that specifically concerned the circadian rhythms of telephone call activity of older adults.

Despite this encouraging pioneering work, key limitations remain to be addressed. In particular, regarding older adults, substantial efforts remain in order to improve the descriptive methodology of the recently completed studies in order to properly model the circadian rhythms of telephone call activity. From a health care monitoring perspective, such efforts are crucial for extracting relevant information from the circadian rhythms of telephone call activity that can be used for completing traditional subjective questionnaires.

Thus, this paper is a secondary analysis of recent analyses [39,40] and is specifically designed to address whether and how the circadian rhythms of social activity observed in the telephone calls of older adults could be modeled. To this end, we analyzed a 12-month data set of outgoing telephone CDRs of 26 adults older than 65 years of age. We designed a statistical learning approach adapted for exploratory analysis. First, Gaussian mixture models (GMMs) were calculated to automatically model each participant's circadian rhythm of telephone call activity. Then, k-means clustering was used for grouping participants into distinct groups depending on the personal characteristics of their GMM curves. The results, the significance, and the limitations of this study are discussed, and perspectives of future research are proposed.

Methods

Data collection, volunteer recruitment, and data preprocessing followed the general principles set in previous investigations of this secondary analysis [39-43] and remain unchanged for consistency.

Data Collection and Volunteer Recruitment

Our data set includes 12 months of outgoing CDRs for 26 volunteers (20 women [77%] and 6 men [23%]; mean age 84 years, SD 4; age range 71-91 years). CDRs provided by the local communication service provider were collected from the personal telephones of the volunteers. Each CDR contained the date, hour, source ID, recipient ID, direction, and duration (seconds) of the call. In addition, volunteers who had several telephones registered with their communication service provider (eg, one or more landlines and/or one or more mobile phones) provided CDRs for all their telephones. Finally, only outgoing calls were considered in this study considering ingoing not resulting from a voluntary activity of the participants. Note that the telephone owners and the telephone contacts remained anonymous. This study and the corresponding experimental protocols were approved by the French Commission for Data Protection and Liberties (Correspondant Informatique et Libertés [CIL] Register France Telecom 2011 n°44) at the time the data collection phase of this project was originally initiated. All methods were performed in accordance with the Commission's regulations, written informed consent was obtained from all participants prior to data collection, and anonymization of participants' data was applied to ensure privacy requirements.

Data Preprocessing

Participants did not all enroll in the survey at the same time, so the dates of inclusion varied. Thus, the CDR data set was filtered to select the time interval of when the greatest number of volunteers were actively participating in the study. The CDR data set was then preprocessed by applying the method described by Saramäki et al [38], namely, by selecting only those participants who used their telephones throughout the entire 12-month observation period, which gave a sample of 21 individuals (see Table 1 for details).

Table 1. Data set structure of call detail records before and after preprocessing.

Participant characteristics	Before preprocessing (N=26)	After preprocessing (N=21)
Number of participants, n (%)		
Female	20 (77)	16 (76)
Male	6 (23)	5 (24)
Age (years)		
Range	71-91	71-91
Mean (SD)	84 (4)	83 (4)
Number of calls, n		
Total outgoing calls	19,198	18,338
Calls by individual, 1st quartile	285	481
Calls by individual, median	590	710
Calls by individual, 3rd quartile	944	1096

Data Analysis

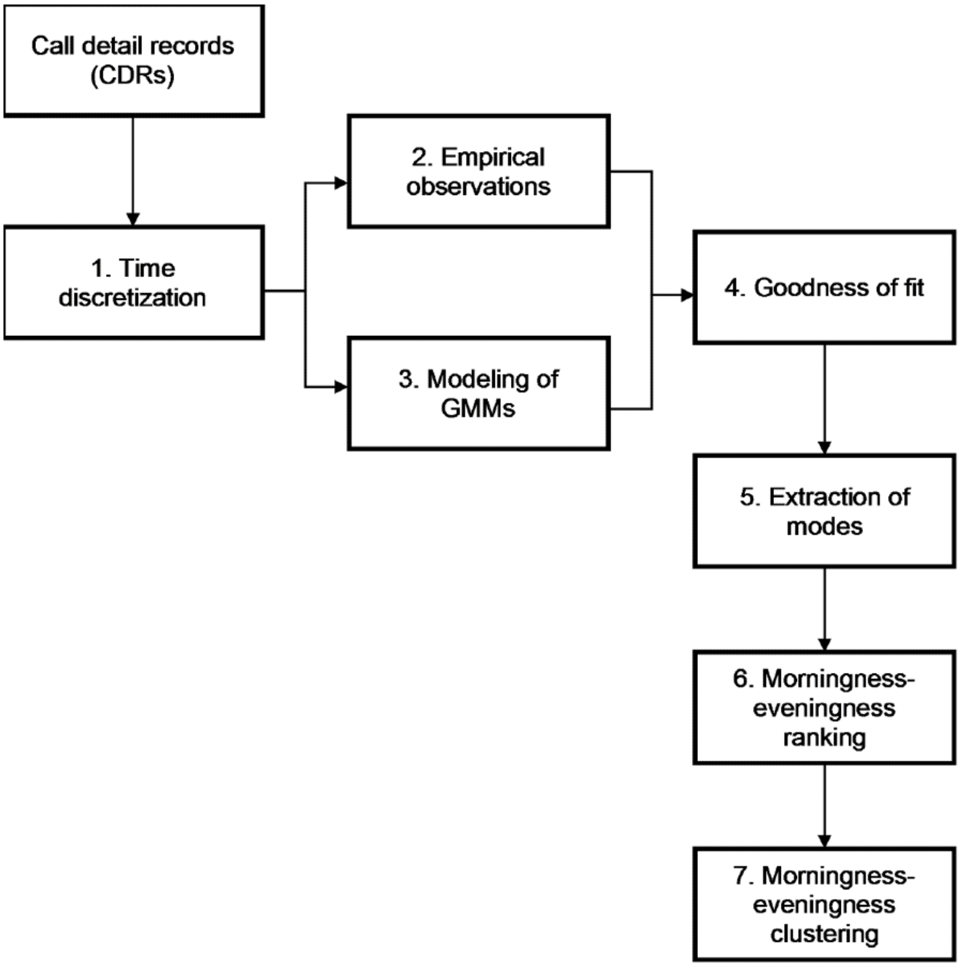
Overview

To identify social interaction patterns in older adults from their CDRs at a daily scale, we developed a statistical learning approach that is divided into two main processes consisting of the following:

1. Modeling daily rhythms of social interactions in older adults from their CDRs by using GMMs [44].
2. Calculating clusters of morningness-eveningness categories stemming from these modeling results by using k-means clustering [45].

These processes were chained together throughout our general data analysis pipeline as discussed in the following seven steps (see Figure 1 for an overview).

Figure 1. Data analysis pipeline. Step 1 corresponds to data preprocessing, Steps 2-4 correspond to circadian rhythm modeling, and Steps 5-7 correspond to social morningness-eveningness clustering. GMM: Gaussian mixture model.



Step 1: Time Discretization

To focus on circadian rhythms of phone call activity, we first discretized the time dimension of CDRs into one unique day of 24 1-hour slots corresponding to the time of day, as previously done in the literature [32,36,39,40,46].

Step 2: Empirical Calculation

From this coarse-grained data set, the empirical circadian rhythm of phone call activity was then obtained by calculating, for each older individual inside of each 1-hour time slot, his or her corresponding frequency of phone calls, as in several studies [32,36,39,40,46]. Formally, if we let N be the number of older participants (N=21), and we let $n_i(t)$ be the number of phone calls of one individual i, with i ranging from 1 to N, at a time t, with t ranging from 0 to 23, then the frequency of phone calls for each individual at a time t is defined as follows:

Step 3: Modeling of GMMs

In parallel with Step 2, a statistical modeling approach caught the distinct periods of phone call activity across the day by calculating, for each older individual, his or her corresponding GMMs from his or her coarse-grained CDRs. In short, this mixture consisted of synthesizing the older individual's circadian rhythm of phone call activity by means of distinct curve components. Since these components stem from the fundamental and well-known Gaussian model, their mean and variance characteristic parameters can be easily calculated and used to further characterize the individual's phone call daily activity. In accordance with the existing literature on the subject [4,47-49], we assumed the existence of two broad periods of activity corresponding to morningness and eveningness in humans, following the diurnal principle, which can be, however, shifted in certain cases toward a potential nocturnal period of activity.

Step 4: Goodness-of-Fit Estimation

The goodness of fit of the estimated GMM with respect to empirical data was assessed for each older individual by using a Kolmogorov-Smirnov comparison test under the null hypothesis H_0—the estimated GMM is similar to the empirical distribution of phone calls across the day—with a *P* value set at .05. Hence, a *P* value greater than .05 represents a significant good fit. This step permits the measurement of the quality of the estimated model with respect to the individual's empirical observations.

Step 5: Extraction of Modes

Next, to focus on the individual's chronotype, the mean parameter of each Gaussian component, which corresponds to an hourly peak of phone call activity, was extracted for each older individual. The duration of each period of activity was also considered by extracting the variance of each period.

Step 6: Morningness-Eveningness Ranking

Since modes were extracted from distinct Gaussian components whose intensities may be different, the individual's morningness-eveningness characteristic was obtained by assigning each older individual to his or her most representative Gaussian component. To this end, the estimated mixing proportion of each Gaussian component was calculated from an expectation minimization approach for each older adult. Next, the individual's modes were ranked together according to these proportions to determine the time of day when the older individual was the most socially active.

Step 7. Morningness-Eveningness Clustering

Eventually, older adults were categorized into distinct groups according to their ranked modes that were calculated above in Step 6. To this end, each older individual was assigned to a point in a 2-dimensional space that corresponded to his or her two periods of phone call activity. We then used k-means clustering to calculate groups of older individuals from the coordinates of their representative points.

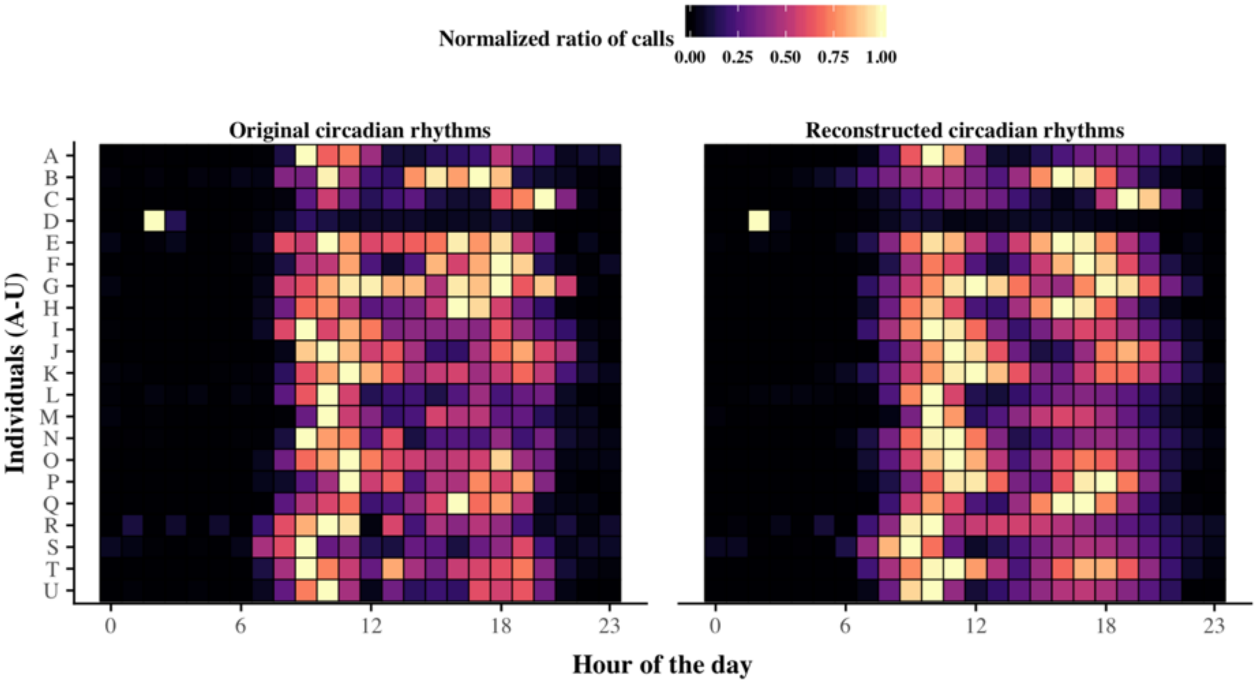
Results

Modeling Circadian Rhythms of Outgoing Phone Call Activity in Older Adults With GMMs

Figure 2 shows, for each older adult, both (1) his or her empirical circadian rhythms of outgoing telephone call activity (see Methods for details) and (2) his or her corresponding GMM (see Methods for details) on two distinct panels. Two observations stand out:

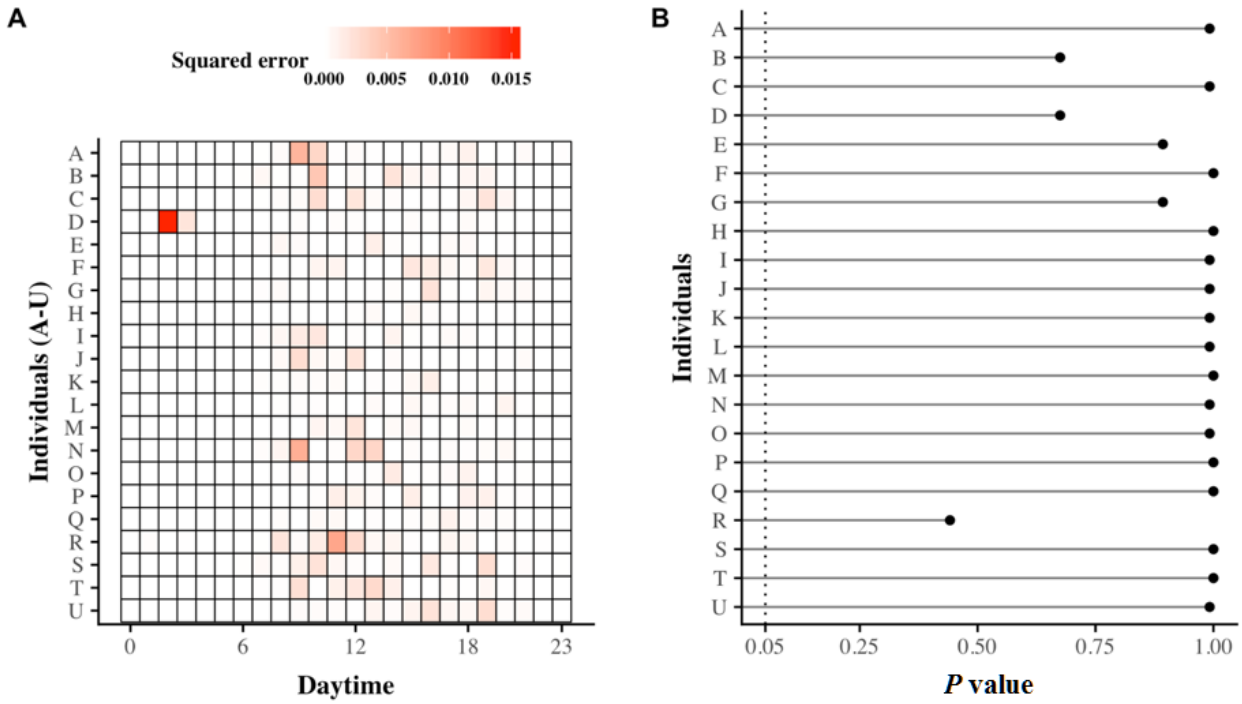
1. On the left-hand panel of Figure 2, the observed empirical circadian rhythms identify the existence of distinct peaks of phone call activity across the day that vary among older individuals. In accordance with previous descriptive results found in Aubourg et al [39], some of them, as with individuals A and I, showed a morning preference for phoning, whereas others, as with individuals C and Q, showed an evening preference. There was also one individual, individual D, who presented a significant shift in his rhythms for phoning, as evidenced by his peak of phone call activity that occurred at night around 2 AM.
2. Comparing empirical data with the models of GMMs for each older individual, we observed on the right-hand panel of Figure 2 the ability of GMMs to highlight periods of high activity and for smoothing periods of weak activity by the smoothing property of the Gaussian curve.

Figure 2. Circadian rhythms of outgoing telephone call activity in older adults and their corresponding Gaussian mixture models (GMMs). These patterns are represented by means of two colored heat maps, where each line corresponds to one older individual and columns correspond to hours of the day. Hence, the outgoing telephone call activity of one individual at a given hour is represented by a colored cell that is all the more bright as the individual increases activity. The left-hand heat map represents the original circadian rhythms of outgoing telephone call activity observed from empirical data, whereas the right-hand heat map represents those modeled by GMMs.



Furthermore, at a statistical level, Figure 3 provides the goodness of fit of the GMMs. In this figure, we observe that the similarity between empirical observations and GMMs cannot be rejected for each of the participants, with a *P* value greater than .05, based on the Kolmogorov-Smirnov comparison test.

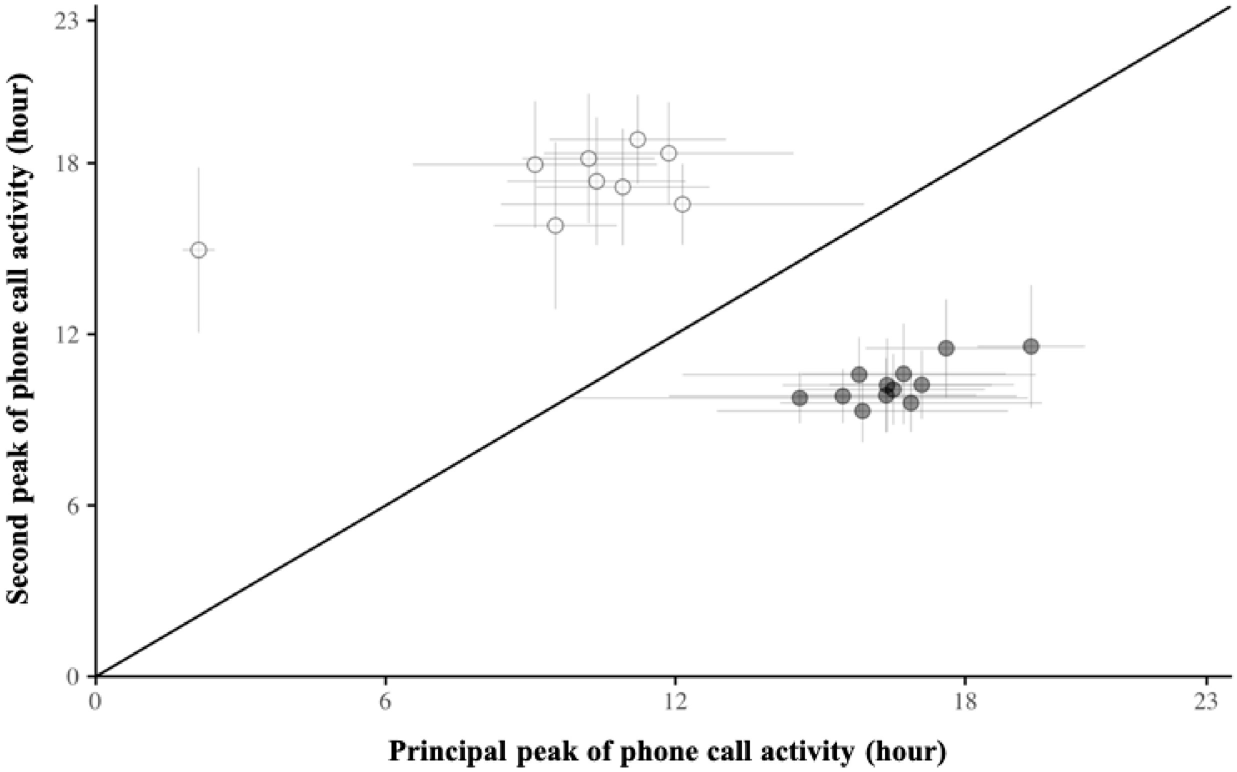
Figure 3. Goodness of fit of Gaussian mixture models (GMMs). Panel A corresponds to a heat map that figures out the squared errors calculated between empirical observations and their corresponding GMMs. Each line corresponds to one older individual, whereas columns correspond to hours of the day. Hence, the squared error between the outgoing telephone call activity and the estimated model at a given hour is represented by a colored cell that is all the more bright as the error increases. Panel B displays the associated goodness of fit by giving *P* values from the Kolmogorov-Smirnov comparison tests. Here, the significance threshold is set at .05 (dashed line).



Social Morningness-Eveningness Is Evidenced in the Telephone Call Activity of Older Adults

The k-means clustering process identified distinct categories of older individuals based on the characteristics of their daily outgoing telephone call activity. Figure 4 shows the k-means clustering results by displaying the two broad periods of telephone call activity of each older individual. This figure highlights the individual nature of social behavior for making telephone calls over the day. In the same way, it stands out that two broad patterns can be conceived in our data set, namely (1) *morningness* individuals have their principal period of telephone call activity in the morning and their second period later in the afternoon or evening and (2) *eveningness* individuals have the exact opposite behavior. Bipartite clustering confirmed this general trend by explicitly exhibiting both of these clusters. Again, we also have to point out that one older adult belonging to the *morningness* cluster, individual D, had an *extreme* behavior whose expression differed from the others, with his first peak of telephone call activity occurring at around 2 AM.

Figure 4. Morningness-eveningness observed in outgoing telephone calls. This figure displays the two broad periods of outgoing telephone call activity. Circles represent peaks of telephone call activity, and each period duration is represented with horizontal and vertical lines. Each circle is colored according to its morningness-eveningness trend given by k-means clustering.



Discussion

Principal Findings

Analysis of daily social interactions by means of CDRs is now recognized as a relevant approach for modeling human activity in an objective, unobtrusive way, which offers new opportunities for improving the field of health monitoring [50,51]. However, in the elderly population, more work is required to confirm the ability of such an approach to model relevant patterns of social interactions at a daily scale. This study is specifically designed to address this issue. To this end, we used a 12-month data set of outgoing CDRs of 26 adults older than 65 years of age. On the whole, our results allowed us to automatically model circadian patterns of outgoing telephone call activity. Interestingly, the results obtained at the individual level suggest the existence of individual hourly preferences for making telephone calls during the day. Despite the evident individual nature of such preferences shown by GMMs, the results also suggest the possibility for assembling these individual observations into broader categories using k-means clustering. In particular, these categories of phone call activity were of two broad types, namely (1) *morningness* older adults that had their main period of telephone call activity in the morning with their second one occurring later, in the afternoon or evening, and (2) *eveningness* older adults that had the exact opposite behavior.

With regard to the field of social physics applied to CDR data sets [52], these results corroborate recent findings about the existence of daily patterns in human social interactions over the telephone, which were exhibited at the population level [53]. Monsivais et al [53] analyzed a 12-month CDR data set containing more than 3 billion calls between 50 million unique identifiers. By aggregating data at the population level, the authors described the existence of an overall bimodality in outgoing telephone call activity that occurred at noon and during the evening periods of around 8 hours. Similarly, they also showed the existence of two periods of around 4 hours of low calling activity occurring (1) at night and (2) after lunch time. Such results permitted the authors to infer the existence of rhythms for representing the resting times of the population in order to study the impact of seasonal and geographical elements on that population. However, the focus on aggregate data from that study [53] does not allow the authors to reach a conclusion regarding the individual rhythms of the studied population because of ecological fallacy, nor within specific populations such as the aging one.

In a complementary way and for the first time to the best of our knowledge, this paper provides evidence at the individual level of the ability of statistical learning approaches to model similar results as those obtained by Monsivais et al [53] within a small cohort of older adults.

Furthermore, two singularities also stand out regarding our results in comparison with those obtained by Monsivais et al [53]. On the one hand, the bimodality exhibited in our older population seemed to appear earlier during the day at midmorning around 10 AM and in the evening around 6 PM. Knowing that the morning transition tends to appear with age, it is plausible that this observed shift, in comparison with the general behavior exhibited in the study by Monsivais et al [53], may reflect such a transitive state and could consequently be a representative element of the aging population [47]. On the other hand, our results also show the existence of nocturnal patterns of telephone call activity within our aging population that could occur in some older individuals. It is plausible that such nocturnal activity could reflect particular disruptions such as sleep disturbances, which are known to occur with age [54].

Hence, the analysis of social interactions in older adults by means of their CDRs does provide evidence about the existence of both social and structural insights about the older individual's daily social life. From a clinical perspective, such insights offer a relevant opportunity to consider modern phone technologies as a new kind of social activity sensor whose captured data can be automatically modeled to identify the fundamental social dynamics in the older individual's life. Social dynamics regarding telephone use could notably be harnessed to detect potential disruptions, such as nocturnal phase shifts or chronotype alterations, or to prevent the occurrence of particular health issues, such as sleep disorders, being presented to the clinician in the form of triggers.

Furthermore, since these pieces of information about social interactions do present the advantage of being collected in a passive, objective, and continuous way, they could be harnessed in complement with other timely subjective methods, such as subjective questionnaires [4,49], but also with innovative digital technological approaches such as *actigraphy*, which already exists for monitoring physical rhythms of the individual [55,56].

Study Limitations and Perspectives

A number of caveats and limitations have to be taken into account. Since this study's analysis was based on a relatively small sample of 26 older individuals, any straightforward generalization of our results to the overall older population should be avoided. Given the sample size of this study, at this stage the interest of our exploratory work is not in providing general knowledge on older adults' social activity on the telephone but, rather, in the ability to gain new insights that deserve to be included for future big data studies on improved data sets. Consequently, whether and how similar results could be observed under different conditions using different sets of data or by implementing alternative statistical methods should naturally be investigated. This could imply working on larger data sets but also, more broadly, leading analyses on distinct populations having their own social characteristics, such as individuals from particular social classes or from different cultures. Given the proportion of women represented in the analyzed data set (76%), whether and how similar works can be replicated with success within populations that have different proportions of male and female participants deserve to be addressed. Furthermore, it could be interesting to focus on specific cohorts suffering from age-related diseases, such as Alzheimer disease [31], Parkinson disease [13], and depression [30,57]. These supplementary works are all the more important insofar as, to the best of our knowledge, this study is the first one that investigates daily patterns of social interactions in an aging population specifically by combining CDR analysis with a statistical learning approach.

Among future perspectives, an essential step that remains to be investigated is whether, and to what extent, the results presented here could be harnessed to enhance already-existing digital frameworks in health research. Such an investigation could consist of developing an adequate complementary framework based on both judiciously used, statistical learning methods and behavioral theory to permit a better understanding of the digital profile of the older individual. In the field of health care monitoring, we do believe that such an investigation could represent a relevant approach for enhancing the interpretability of results and the efficiency of modeling approaches relying on telephone call activity.

Conclusions

Findings from this study support the potential of phone technologies and statistical learning approaches for automatically providing personalized and precise insights on the social rhythms of telephone call activity of older individuals. Such insights can highlight particular individual habits related to being active on the telephone according to the time of day. Furthermore, these individual pieces of information can be automatically analyzed by statistical learning approaches at the group level in order to provide more general information on the observed cohort, such as subtrends for clusters of older individuals, for instance, who are more socially active on the telephone in the morning or in the evening. Futures studies could integrate such digital insights with other sources of data to complete the assessment of the circadian rhythms of activity in the elderly population. In particular, big data studies should be implemented with large cohorts in order to investigate to what extent our study's results that were obtained from our sample of older participants at the group level can be transformed into broad actionable knowledge for the older population.

Acknowledgments

We thank the reviewers for critically reading the manuscript and suggesting substantial improvements.

Authors' Contributions

TA, NV, and JD designed the study. TA performed the data analysis, prepared the figures, and wrote the main manuscript text under the supervision of NV and JD. NV contributed to data acquisition. NV and JD substantively revised the manuscript text. NV contributed to the interpretation of the data.

Conflicts of Interest

None declared.

References

1. Herzog ED. Neurons and networks in daily rhythms. Nat Rev Neurosci 2007 Oct;8(10):790-802. [doi: 10.1038/nrn2215] [Medline: 17882255]
2. Abbott SM, Knutson KL, Zee PC. Health implications of sleep and circadian rhythm research in 2017. Lancet Neurol 2018 Jan;17(1):17-18. [doi: 10.1016/s1474-4422(17)30410-6]
3. Refinetti R, Menaker M. The circadian rhythm of body temperature. Physiol Behav 1992 Mar;51(3):613-637. [doi: 10.1016/0031-9384(92)90188-8]
4. Monk TH, Flaherty JF, Frank E, Hoskinson K, Kupfer DJ. The Social Rhythm Metric. An instrument to quantify the daily rhythms of life. J Nerv Ment Dis 1990 Feb;178(2):120-126. [doi: 10.1097/00005053-199002000-00007] [Medline: 2299336]
5. Goldbeter A. Au Cœur des Rythmes du Vivant: La Vie Oscillatoire. Paris, France: Odile Jacob; 2018.
6. Hood S, Amir S. The aging clock: Circadian rhythms and later life. J Clin Invest 2017;127:446 [FREE Full text] [doi: 10.1172/jci90328]
7. Steponenaite A, Biello SM, Lall GS. Aging clocks: Disrupted circadian rhythms. Aging (Albany NY) 2018 Nov 13;10(11):3065-3066 [FREE Full text] [doi: 10.18632/aging.101642] [Medline: 30425185]
8. Zhao J, Warman GR, Cheeseman JF. The functional changes of the circadian system organization in aging. Ageing Res Rev 2019 Jul;52:64-71. [doi: 10.1016/j.arr.2019.04.006] [Medline: 31048031]
9. Duffy JF, Zitting K, Chinoy ED. Aging and circadian rhythms. Sleep Med Clin 2015 Dec;10(4):423-434 [FREE Full text] [doi: 10.1016/j.jsmc.2015.08.002] [Medline: 26568120]
10. Chen C, Logan RW, Ma T, Lewis DA, Tseng GC, Sibille E, et al. Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. Proc Natl Acad Sci U S A 2016 Jan 05;113(1):206-211 [FREE Full text] [doi: 10.1073/pnas.1508249112] [Medline: 26699485]
11. Maury E. Off the clock: From circadian disruption to metabolic disease. Int J Mol Sci 2019 Mar 30;20(7):1597 [FREE Full text] [doi: 10.3390/ijms20071597] [Medline: 30935034]
12. Leng Y, Musiek ES, Hu K, Cappuccio FP, Yaffe K. Association between circadian rhythms and neurodegenerative diseases. Lancet Neurol 2019 Mar;18(3):307-318. [doi: 10.1016/s1474-4422(18)30461-7]
13. Stewart J, Bachman G, Cooper C, Liu L, Ancoli-Israel S, Alibiglou L. Circadian dysfunction and fluctuations in gait initiation impairment in Parkinson's disease. Exp Brain Res 2018 Mar;236(3):655-664. [doi: 10.1007/s00221-017-5163-5] [Medline: 29294143]
14. Musiek ES, Holtzman DM. Mechanisms linking circadian clocks, sleep, and neurodegeneration. Science 2016 Nov 25;354(6315):1004-1008 [FREE Full text] [doi: 10.1126/science.aah4968] [Medline: 27885006]
15. Hanefeld M, Engelmann K, Appelt D, Sandner D, Weigmann I, Ganz X, et al. Intra-individual variability and circadian rhythm of vascular endothelial growth factors in subjects with normal glucose tolerance and type 2 diabetes. PLoS One 2017;12(10):e0184234 [FREE Full text] [doi: 10.1371/journal.pone.0184234] [Medline: 28991900]
16. Cornelissen G, Otsuka K. Chronobiology of aging: A mini-review. Gerontology 2017;63(2):118-128 [FREE Full text] [doi: 10.1159/000450945] [Medline: 27771728]
17. Otsuka K, Cornelissen G, Halberg F. Chronomics and Continuous Ambulatory Blood Pressure Monitoring. Vascular Chronomics: From 7-Day/24-Hour to Lifelong Monitoring. Tokyo, Japan: Springer Japan; 2016.
18. Tazawa Y, Wada M, Mitsukura Y, Takamiya A, Kitazawa M, Yoshimura M, et al. Actigraphy for evaluation of mood disorders: A systematic review and meta-analysis. J Affect Disord 2019 Jun 15;253:257-269. [doi: 10.1016/j.jad.2019.04.087] [Medline: 31060012]
19. Schwab KE, Ronish B, Needham DM, To AQ, Martin JL, Kamdar BB. Actigraphy to evaluate sleep in the intensive care unit. A systematic review. Ann Am Thorac Soc 2018 Sep;15(9):1075-1082. [doi: 10.1513/annalsats.201801-004oc]
20. Koolhaas CM, Dhana K, van Rooij FJ, Kocevska D, Hofman A, Franco OH, et al. Sedentary time assessed by actigraphy and mortality: The Rotterdam Study. Prev Med 2017 Feb;95:59-65. [doi: 10.1016/j.ypmed.2016.11.021] [Medline: 27939262]
21. Dean GE, Ziegler P, Chen H, Steinbrenner LM, Dickerson SS. Trajectory of insomnia symptoms in older adults with lung cancer: Using mixed methods. Support Care Cancer 2019 Jun;27(6):2255-2263. [doi: 10.1007/s00520-018-4488-3] [Medline: 30324283]
22. Kikkert LHJ, Vuillerme N, van Campen JP, Appels BA, Hortobágyi T, Lamoth CJC. The relationship between gait dynamics and future cognitive decline: A prospective pilot study in geriatric patients. Int Psychogeriatr 2017 Dec 10;30(9):1301-1309. [doi: 10.1017/s1041610217002770]
23. Yan H, Xu LD, Bi Z, Pang Z, Zhang J, Chen Y. An emerging technology – Wearable wireless sensor networks with applications in human health condition monitoring. J Manag Anal 2015 Apr 13;2(2):121-137. [doi: 10.1080/23270012.2015.1029550]
24. Grandin LD, Alloy LB, Abramson LY. The social zeitgeber theory, circadian rhythms, and mood disorders: Review and evaluation. Clin Psychol Rev 2006 Oct;26(6):679-694. [doi: 10.1016/j.cpr.2006.07.001] [Medline: 16904251]

25. Froy O, Miskin R. The interrelations among feeding, circadian rhythms and ageing. Prog Neurobiol 2007 Jun;82(3):142-150. [doi: 10.1016/j.pneurobio.2007.03.002] [Medline: 17482337]
26. Weinmann T, Vetter C, Karch S, Nowak D, Radon K. Shift work and cognitive impairment in later life - Results of a cross-sectional pilot study testing the feasibility of a large-scale epidemiologic investigation. BMC Public Health 2018 Nov 14;18(1):1256 [FREE Full text] [doi: 10.1186/s12889-018-6171-5] [Medline: 30428871]
27. Lowenthal MF. Social isolation and mental illness in old age. Am Sociol Rev 1964 Feb;29(1):54. [doi: 10.2307/2094641]
28. Malone S, Mendoza M, Patterson F. Social jetlag, circadian disruption, and cardiometabolic disease risk. Sleep Health 2019:227-240. [doi: 10.1016/b978-0-12-815373-4.00018-6]
29. Quante M, Mariani S, Weng J, Marinac CR, Kaplan ER, Rueschman M, et al. Zeitgebers and their association with rest-activity patterns. Chronobiol Int 2019 Feb;36(2):203-213 [FREE Full text] [doi: 10.1080/07420528.2018.1527347] [Medline: 30365354]
30. Aung MN, Moolphate S, Aung TNN, Kantonyoo C, Khamchai S, Wannakrairot P. The social network index and its relation to later-life depression among the elderly aged ≥80 years in Northern Thailand. Clin Interv Aging 2016 Aug;Volume 11:1067-1074. [doi: 10.2147/cia.s108974]
31. Musiek ES, Bhimasani M, Zangrilli MA, Morris JC, Holtzman DM, Ju YS. Circadian rest-activity pattern changes in aging and preclinical Alzheimer disease. JAMA Neurol 2018 May 01;75(5):582-590 [FREE Full text] [doi: 10.1001/jamaneurol.2017.4719] [Medline: 29379963]
32. Aledavood T, López E, Roberts SGB, Reed-Tsochas F, Moro E, Dunbar RIM, et al. Daily rhythms in mobile telephone communication. PLoS One 2015;10(9):e0138098 [FREE Full text] [doi: 10.1371/journal.pone.0138098] [Medline: 26390215]
33. Althoff T, Horvitz E, White R, Zeitzer J. Harnessing the web for population-scale physiological sensing: A case study of sleep and performance. In: Proceedings of the 26th International World Wide Web Conference. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2017 Presented at: 26th International World Wide Web Conference; April 3-7, 2017; Perth, Australia p. 113-122. [doi: 10.1145/3038912.3052637]
34. Murnane E, Abdullah S, Matthews M, Choudhury T, Gay G. Social (media) jet lag: How usage of social technology can modulate-reflect circadian rhythms. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive-Ubiquitous Computing. New York, NY: Association for Computing Machinery; 2015 Presented at: 2015 ACM International Joint Conference on Pervasive-Ubiquitous Computing; September 7-11, 2015; Osaka, Japan p. 843-854. [doi: 10.1145/2750858.2807522]
35. Blondel VD, Decuyper A, Krings G. A survey of results on mobile phone datasets analysis. EPJ Data Sci 2015 Aug 5;4:1-55 [FREE Full text] [doi: 10.1140/epjds/s13688-015-0046-0]
36. Aledavood T, Lehmann S, Saramäki J. Digital daily cycles of individuals. Front Phys 2015 Oct 07;3:1-7 [FREE Full text] [doi: 10.3389/fphy.2015.00073]
37. Aledavood T, Lehmann S, Saramäki J. Social network differences of chronotypes identified from mobile phone data. EPJ Data Sci 2018 Oct 24;7:1-13 [FREE Full text] [doi: 10.1140/epjds/s13688-018-0174-4]
38. Saramäki J, Leicht EA, López E, Roberts SGB, Reed-Tsochas F, Dunbar RIM. Persistence of social signatures in human communication. Proc Natl Acad Sci U S A 2014 Jan 21;111(3):942-947 [FREE Full text] [doi: 10.1073/pnas.1308540110] [Medline: 24395777]
39. Aubourg T, Demongeot J, Provost H, Vuillerme N. Circadian rhythms in the telephone calls of older adults: Observational descriptive study. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e12452 [FREE Full text] [doi: 10.2196/12452] [Medline: 32130156]
40. Aubourg T, Demongeot J, Provost H, Vuillerme N. How to measure circadian rhythms of activity and their disruptions in humans using passive and unobtrusive capture of phone call activity. In: Proceedings of the 17th World Congress of Medical and Health Informatics (MEDINFO 2019). 2019 Presented at: 17th World Congress of Medical and Health Informatics (MEDINFO 2019); August 26-30, 2019; Lyon, France p. 1631-1632. [doi: 10.3233/SHTI190569]
41. Aubourg T, Demongeot J, Renard F, Provost H, Vuillerme N. Association between social asymmetry and depression in older adults: A phone call detail records analysis. Sci Rep 2019 Sep 18;9(1):13524 [FREE Full text] [doi: 10.1038/s41598-019-49723-8] [Medline: 31534178]
42. Aubourg T, Demongeot J, Provost H, Vuillerme N. Exploitation of outgoing and incoming telephone calls in the context of circadian rhythms of social activity among elderly people: Observational descriptive study. JMIR Mhealth Uhealth 2020 Nov 26;8(11):e13535 [FREE Full text] [doi: 10.2196/13535] [Medline: 33242018]
43. Aubourg T, Demongeot J, Vuillerme N. Novel statistical approach for assessing the persistence of the circadian rhythms of social activity from telephone call detail records in older adults. Sci Rep 2020 Dec 08;10(1):21464 [FREE Full text] [doi: 10.1038/s41598-020-77795-4] [Medline: 33293551]
44. Celeux G, Govaert G. Fuzzy clustering and mixture models. In: Proceedings of the 11th Symposium on Computational Statistics (Compstat). Heidelberg, Germany: Physica-Verlag GmbH & Co; 1994 Presented at: 11th Symposium on Computational Statistics (Compstat); August 22-26, 1994; Vienna, Austria p. 154-159. [doi: 10.1007/978-3-642-52463-9_15]
45. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognit Lett 2010 Jun;31(8):651-666. [doi: 10.1016/j.patrec.2009.09.011]

46. Aledavood T, López E, Roberts S, Reed-Tsochas F, Moro E, Dunbar R. Channel-specific daily patterns in mobile phone communication. In: Proceedings of the European Conference on Complex Systems. Cham, Switzerland: Springer International Publishing; 2014 Presented at: European Conference on Complex Systems; September 22-26, 2014; Lucca, Italy p. 209-218. [doi: 10.1007/978-3-319-29228-1_18]
47. Roenneberg T. What is chronotype? Sleep Biol Rhythms 2012;10:75-76. [doi: 10.1111/j.1479-8425.2012.00541.x]
48. Monk TH, Buysse DJ. Chronotype, bed timing and total sleep time in seniors. Chronobiol Int 2014 Jun;31(5):655-659 [FREE Full text] [doi: 10.3109/07420528.2014.885981] [Medline: 24517139]
49. Roenneberg T, Wirz-Justice A, Merrow M. Life between clocks: Daily temporal patterns of human chronotypes. J Biol Rhythms 2003 Feb;18(1):80-90. [doi: 10.1177/0748730402239679] [Medline: 12568247]
50. Bhugra D, Tasman A, Pathare S, Priebe S, Smith S, Torous J, et al. The WPA-Lancet Psychiatry Commission on the future of psychiatry. Lancet Psychiatry 2017 Oct;4(10):775-818. [doi: 10.1016/S2215-0366(17)30333-4] [Medline: 28946952]
51. Russ TC, Woelbert E, Davis KAS, Hafferty JD, Ibrahim Z, Inkster B, MQ Data Science Group. How data science can advance mental health research. Nat Hum Behav 2019 Jan;3(1):24-32. [doi: 10.1038/s41562-018-0470-9] [Medline: 30932051]
52. Bhattacharya K, Kaski K. Social physics: Uncovering human behaviour from communication. Adv Phys X 2018 Nov 14;4(1):96-124 [FREE Full text] [doi: 10.1080/23746149.2018.1527723]
53. Monsivais D, Bhattacharya K, Ghosh A, Dunbar RIM, Kaski K. Seasonal and geographical impact on human resting periods. Sci Rep 2017 Sep 06;7(1):10717 [FREE Full text] [doi: 10.1038/s41598-017-11125-z] [Medline: 28878235]
54. Ancoli-Israel S. Insomnia in the elderly: A review for the primary care practitioner. Sleep 2000 Feb 01;23 Suppl 1:S23-S30; discussion S36-S38. [Medline: 10755805]
55. Benard V, Etain B, Vaiva G, Boudebesse C, Yeim S, Benizri C, et al. Sleep and circadian rhythms as possible trait markers of suicide attempt in bipolar disorders: An actigraphy study. J Affect Disord 2019 Feb 01;244:1-8. [doi: 10.1016/j.jad.2018.09.054] [Medline: 30290235]
56. Ho S, Mohtadi A, Daud K, Leonards U, Handy TC. Using smartphone accelerometry to assess the relationship between cognitive load and gait dynamics during outdoor walking. Sci Rep 2019 Feb 28;9(1):3119 [FREE Full text] [doi: 10.1038/s41598-019-39718-w] [Medline: 30816292]
57. Domènech-Abella J, Lara E, Rubio-Valera M, Olaya B, Moneta MV, Rico-Uribe LA, et al. Loneliness and depression in the elderly: The role of social network. Soc Psychiatry Psychiatr Epidemiol 2017 Apr;52(4):381-390. [doi: 10.1007/s00127-017-1339-3] [Medline: 28154893]

Abbreviations

CDR: call detail record
CIL: Correspondant Informatique et Libertés
GMM: Gaussian mixture model

Edited by R Kukafka; submitted 09.07.20; peer-reviewed by I Mircheva, R De Carvalho; comments to author 17.08.20; revised version received 02.09.20; accepted 07.09.20; published 08.01.21.

Please cite as:
Aubourg T, Demongeot J, Vuillerme N
Gaining Insights Into the Estimation of the Circadian Rhythms of Social Activity in Older Adults From Their Telephone Call Activity With Statistical Learning: Observational Study
J Med Internet Res 2021;23(1):e22339
URL: http://www.jmir.org/2021/1/e22339/
doi:10.2196/22339
PMID:33416502

Original Paper

Sharing Patient-Controlled Real-World Data Through the Application of the Theory of Commons: Action Research Case Study

Andreas Hager[1], LLM; Staffan Lindblad[1], MD, PhD; Mats Brommels[2], MD, PhD; Stina Salomonsson[3], PhD; Carolina Wannheden[2], PhD

[1]Upstream Dream AB, Bromma, Sweden
[2]Medical Management Centre, Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden
[3]Center for Observational and Real World Evidence, Merck Sharp and Dohme, Stockholm, Sweden

Corresponding Author:
Carolina Wannheden, PhD
Medical Management Centre
Department of Learning, Informatics, Management and Ethics (LIME)
Karolinska Institutet
LIME, 4th floor
Tomtebodavägen 18A
Stockholm, 171 77
Sweden
Phone: 46 8 524 83934
Email: carolina.wannheden@ki.se

Abstract

Background: Technological advances have radically changed the opportunities for individuals with chronic conditions to practice self-care and to coproduce health care and research. Digital technologies enable patients to perform tasks traditionally carried out by health care professionals in a more convenient way, at lower costs, and without compromising quality. Patients may also share real-world data with other stakeholders to promote individual and population health. However, there is a need for legal frameworks that enable patient privacy and control in such sharing of real-world data. We believe that this need could be met by the conceptualization of patient-controlled real-world data as knowledge commons, which is a resource shared by a group of people.

Objective: This study aimed to propose a conceptual model that describes how patient-controlled real-world data can be shared effectively in chronic care management, in a way that supports individual and population health, while respecting personal data privacy and control.

Methods: An action research approach was used to develop a solution to enable patients, in a self-determined way, to share patient-controlled data to other settings. We chose the context of cystic fibrosis (CF) care in Sweden, where coproduction between patients, their families, and health care professionals is critical in the introduction of new drugs. The first author, who is a lawyer and parent of children with CF, was a driver in the change process. All coauthors collaborated in the analysis. We collected primary and secondary data reflecting changes during the time period from 2012 to 2020, and performed a qualitative content analysis guided by the knowledge commons framework.

Results: Through a series of changes, a national system for enabling patients to share patient-controlled real-world data to different stakeholders in CF care was implemented. The case analysis resulted in a conceptual model consisting of the following three knowledge commons arenas that contributed to patient-controlled real-world data collection, use, and sharing: (1) patient world arena involving the private sphere of patients and families; (2) clinical microsystem arena involving the professional sphere at frontline health care clinics; and (3) round table arena involving multiple stakeholders from different settings. Based on the specification of property rights, as presented in our model, the patient can keep control over personal health information and may grant use rights to other stakeholders.

Conclusions: Health information exchanges for sharing patient-controlled real-world data are pivotal to enable patients, health care professionals, health care funders, researchers, authorities, and the industry to coproduce high-quality care and to introduce and follow-up novel health technologies. Our model proposes how technical and legal structures that protect the integrity and self-determination of patients can be implemented, which may be applicable in other chronic care settings as well.

(J Med Internet Res 2021;23(1):e16842) doi:10.2196/16842

KEYWORDS

knowledge commons; learning networks; patient and family centered care; eHealth

Introduction

Moving From Self-Care and Health Care to Co-Care

In high-income countries, the top 1% of high-cost patients, predominantly patients with complex chronic care needs, account for almost one-fourth (24%) of health care expenditure [1]. Over half of high-cost patients are under the age of 65 years [1]. As the population with chronic illness grows, health care is under increasing pressure to use available resources efficiently, without compromising quality of care. In recent years, the necessity of stronger citizen involvement in health care service delivery has been emphasized using several related concepts, such as patient and public involvement [2], shared decision making [3], patient- and person-centered care [4,5], and coproduction [6].

Ostrom defines coproduction as “the process through which inputs used to produce goods or services are contributed by individuals who are not in the same organization” [7]. Von Thiele Schwarz [8] uses the term “co-care” to describe coproduction between patients and other actors involved in their care, such as health care professionals, and emphasizes the usefulness of digital technologies in facilitating the exchange of knowledge and experience between different actors. Digital technologies enable patients to perform tasks traditionally carried out by health care staff in a more convenient way, at lower costs, and without compromising quality [9]. For example, people measure their own blood pressure, adjust their insulin dose based on their own glucose monitoring, and perform intravenous antibiotic treatments at home, without the support or direct monitoring of health care staff.

Learning Health Care Systems as Knowledge Commons

With technology literally in their back pockets, individuals produce large personal health information repositories that can be shared electronically [10]. In the context of this study, we use the term *patient-controlled real-world data* to describe personal health-related data that are controlled by patients. Patient control implies that patients, in line with the General Data Protection Regulation (GDPR) [11], determine the purposes and means of the processing of their personal data, which does not prevent other parties from also being authorized as controllers, by the patient or by law. If adequately utilized, patient-controlled real-world data collected by patients and informal caregivers, in combination with health care collected data, can fuel the development of learning health care systems [12,13]. The information and knowledge that are shared in such collaborative systems can be conceptualized as *commons* [14]. The concept of commons refers to resources that are shared by a group of people and has its roots in the study of shared natural resources (eg, land and water) that are subject to social dilemmas such as competition, freeriding, and overharvesting [15]. In the 1990s, the concept gained application in the study of distributed digital information and knowledge, which Hess and Ostrom describe as knowledge commons [15].

In knowledge commons composed of personal health data, two social dilemmas in particular have been highlighted, namely disempowerment and enclosure [16]. Disempowerment of the individual in relation to the information industry has been described as a problem of unsustainable data practices that can also be labeled as a “privacy harm” resulting from uncontrolled data analytics [16]. As we “*bleed* data when moving a mouse cursor,” large industry actors like search engines gain power over us, which they can use to nudge us into certain behaviors (eg, by directed advertisements) [16]. The enclosure of data refers to the exclusive access to data by certain actors in the information industry, while not allowing others to profit from it, which is a type of appropriation problem [16]. The individual patient is the only source of real-time observations and experiences that span a lifetime of chronic illness. Several stakeholders have large resources and a mission to collect and act on patient-reported real-world data (eg, health care funders, national authorities, and the industry). Disempowerment and enclosure are prevalent social dilemmas in the current system of data flow between stakeholders, which hampers the effective development, introduction, and evaluation of new health technologies.

The knowledge commons is a useful framework to identify such problems based on how information is produced and used by different stakeholders, and to guide the development of suitable solutions to address these, for example, by the definition of social rules and legal mechanisms that enable individuals to effectively share ownership and control of resources [17]. To the best of our knowledge, strategies to explore and effectively manage such dilemmas in the health care and life science industries are limited. Therefore, in this study, we use the knowledge commons framework to explore how to develop adequate technical and legal structures for sharing patient-controlled real-world data.

Aim

This study aimed to propose a conceptual model that describes how patient-controlled real-world data can be shared effectively in chronic care management, in a way that supports individual and population health, while respecting personal data privacy and control.

Methods

Study Design

We used an action research approach that is suitable when the purpose is to bring about change in real-world environments [18]. Action research is focused on solving a specific problem in a specific context. Contribution to both practice and research is achieved through the combined expertise of practitioners and researchers, which mandates a collaborative research approach [19]. While no specific method of data collection is specified, a variety of generally qualitative methods may be applied, including journal keeping, document collection, and case studies [18]. In the following subsections, we first describe the chronic care context that was chosen to address the aim of this study. Thereafter, we describe the action research process and the data collection and analysis methods that were used.

Chronic Care Context

We selected the context of cystic fibrosis (CF) care in Sweden. This context was chosen because CF care and its improvement are strongly dependent on coproduction between patients, their families, and health care [20,21]. CF is a life-limiting recessive genetic disorder that is usually diagnosed within the first few years of life, affects several organs, and leads to chronic infection and inflammation in the airways [21]. Treatment involves drug therapy, daily self-management (eg, airway clearance and physical exercise), and support by multidisciplinary care teams [22]. The recent development of disease-modifying drug therapies has substantially improved the prospect of effective treatment for CF [23]. The availability and use of solutions that enable patients and their families to share patient-controlled real-world data are critical to support the introduction and follow-up of these new therapies. In Sweden, there are approximately 700 CF patients, and about 95% of them are included in the national CF quality registry where care process and outcome data are reported for benchmarking and quality improvement [24].

Action Research Process

The main change agent in the action research process was the first author (AH) who is a lawyer and parent of children with CF, as well as a driver in the development of mobile patient support systems (PSSs) to support patients and their families in self-care and communication with health care. To adequately explore the development of a new infrastructure for sharing of patient-controlled real-world data in the context of CF care in Sweden, the researchers collaborated with AH, combining their contextual and theoretical knowledge to contribute to problem-solving and knowledge generation. Action research is cyclical by nature, which means that changes are continually evaluated, leading to incremental improvements. While the change process itself is not the focus of this study, we analyzed the resulting changes that enabled patients to self-determinedly share patient-controlled real-world data to different stakeholders in CF care. The action research process was guided by the knowledge commons framework.

Knowledge Commons Framework

The knowledge commons framework defines concepts that are central to understanding knowledge as a shared resource. A central concept is the *action arena*, which consists of the individuals and organizations who make decisions that affect patterns of interactions and outcomes in a knowledge commons [25]. To understand the complex nature of information and knowledge sharing, the knowledge commons framework makes a distinction between *ideas*, *artifacts*, and *facilities* [26]. An *idea* is an intangible representation of information in an individual's mind. For example, someone's experiences or thoughts that have not yet been translated into natural language. An *artifact* is a discreet and nameable representation of an idea or a collection of ideas. For example, an electronic health record is a digital collection of ideas about an individual's medical history. Digital artifacts can be used concurrently by multiple users and are stored in information facilities. A *facility* consists of the software and hardware that stores artifacts and makes these available to users. The distinction between ideas, artifacts, and facilities is important for the specification of legal structures in commons. To ensure sustainable use of a shared resource, different sets of rules need to be applied. One set of rules regulates access to the facilities of commons and another set of rules regulates the contribution and withdrawal of artifacts to and from the commons. Seven types of property rights have been identified as relevant in rule setting in digital information commons [25]. These concern the right to access a physical area (*access right*); the right to contribute to content (*contribution right*), extract information (*extraction right*), and remove information (*removal right*); the right to regulate use patterns or make changes to a facility, for example, by adding new functionalities to an application (*management/participation right*); the right to determine access, extraction, and removal rights (*exclusion right*); and the right to sell or lease extraction, management/participation, and exclusion rights (*alienation right*).

Data Collection and Analysis

We collected primary data in the form of formal documents (project plans, meeting notes, technical specifications, contracts, and user manuals) and secondary data in the form of scientific publications [27,28], a master's thesis [29,30], and a conference abstract [31] reflecting the design, development, implementation, and adaptation of changes during the time period from 2012 to 2020. These data were corroborated by the personal experience of AH who was a driver in the change process.

We performed a qualitative analysis resembling a directed content analysis [32], guided by the knowledge commons framework. Two authors (AH and SL) read through the collected documents and identified text that matched any of the knowledge commons concepts described above, that is, descriptions of action arenas, facilities, artifacts, and ideas, as well as descriptions of rules or property rights that were applied to protect personal privacy and control. Text that matched these concepts was extracted, labeled, and categorized. The relations between concepts were explored and summarized into models that visualize these relationships. The models were refined in discussions among all the coauthors, resulting in the conceptual model presented here.

Results

Overview

The first section briefly summarizes the main developments that resulted from the action research project and allowed patients and families to share patient-controlled real-world data with different stakeholders in CF care. Where applicable, references to relevant publications that report these changes are provided in parentheses. The second section presents our analysis of the implemented changes, which is presented as a conceptual model consisting of the different action arenas that we identified and the patterns of patient-controlled real-world data flow between them.

Main Developments

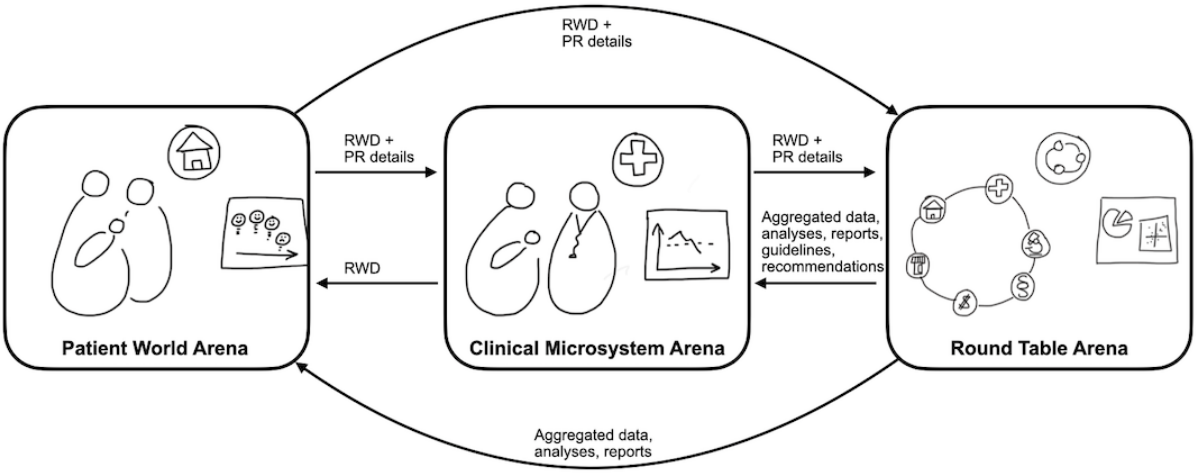
In 2014, a mobile patient-facing application, Genia PSS, was introduced at one of the four national CF centers [27]. The Genia PSS was designed to help patients and their families keep track of patients' health and self-care activities, and to facilitate communication, relationship building, and shared decision making together with their care team. Using the Genia PSS, patients could track and share their own health observations with their care teams. A health information exchange was developed for enabling import of shared data from the Genia PSS into the national CF quality registry.

In late 2015, a new and very costly disease-modifying CF combination therapy, lumacaftor/ivacaftor, was approved for market introduction in Sweden [33]. Not all patients eligible for therapy are responders, and it requires structured follow-up and evaluation of treatment effectiveness, following the national managed introduction process [34]. In 2016, the Sweden CF Coalition was formed as a collaborative learning network involving a variety of stakeholders, comprising representatives from all four national CF Centers, the National Cystic Fibrosis Association, the CF Working Group of the Swedish Society of Medicine, the National Quality Registry Development Group, and the Genia PSS Development Group. Author AH had a role as secretary, and SL served as an advisor. The CF Coalition defined their purpose as follows: "[to make] it easier for everyone – persons with CF, families, clinicians, researchers, and others – to work together to improve health, care, and cost of illness for people with CF" (Sweden Coalition Charter, May 21, 2016). In 2017, they formed a Coordination Group to take on the specific task to implement a national system for the orderly introduction and follow-up of lumacaftor/ivacaftor, using patient-collected and health care–collected real-world data [31]. Through the establishment of a multistakeholder learning network and implementation of new health information exchange services to the national quality registry, the Genia PSS enabled the use of patient-controlled real-world data to support quality improvement, research, and the orderly introduction and follow-up of new CF therapies. In 2018, when the treatment follow-up routines were in place, lumacaftor/ivacaftor was approved for subsidy in Sweden and introduced on the market. At the time of writing, each of the four national CF centers was connected to the Genia PSS, which, in October 2020, had about 240 active users with CF in Sweden.

Three Arenas for Collaboration

The analysis resulted in a conceptual model describing how patient-controlled real-world data can be used and shared across the following three action arenas in CF care: patient world arena, clinical microsystem arena, and round table arena (Figure 1). In the following sections, we describe the three arenas and the property rights that were specified to manage patient-controlled real-world data.

Figure 1. Three central arenas in the collection, use, and sharing of patient-controlled real-world data (RWD). PR details: specification of property rights that translate into use restrictions.



Patient World Arena

The *patient world arena* comprises the private sphere of patients and families living with CF. Using the terminology introduced by Kleinman, the actors in this arena cope with *illness*, *illness problems*, and *illness behavior* [35]. The patient and family together observe and experience health and well-being in the presence of *illness*, as well as the principal difficulties that symptoms and disabilities create for the patient (*illness problems*). They make judgements about how best to cope with the distress and with the practical problems in daily living, including initiating treatment, handling self-management, and deciding when to seek care from professionals and others (*illness behavior*).

Collection and Use of Patient World Artifacts

The Genia PSS enables actors in the patient world arena to capture nonphysical ideas of personal importance about illness, illness problems, and illness behavior, and generate physical artifacts that can be communicated and shared. Artifacts may represent assessments (eg, perceived health benefit from antibiotic treatment), memos (eg, a picture of a run in the forest), or narratives (eg, a treatment diary). The use of additional facilities, such as a blood glucose monitor, a pulse sensor, a scale, and a spirometer, may be necessary to capture health parameters of importance (eg, blood glucose levels, pulse, body weight, and lung function). Personal health data can also be imported for private use in the patient world arena from facilities used in other settings, such as electronic health record (EHR) and pharmacy systems. Collected and imported patient-controlled real-world data are stored in a personal database that can be made available via interfaces with patient-determined members of the patient world arena, such as family members and close friends.

Sharing of Patient World Artifacts With Other Arenas

Health information exchanges were implemented to support sharing of patient-controlled real-world data with other arenas, for example, with health care professionals prior to visits [28]. The sharing of artifacts with other facilities requires mutual agreement and adherence to information standards and terminologies (semantic interoperability), as well as contractual obligations and data protection legislation (legal interoperability). The Genia PSS has implemented a *previsit form* data structure to enable the sharing of patient-controlled real-world data with the clinics via EHR systems and the national quality registry.

Property Rights

The patient world arena represents a private domain within which patient-controlled real-world data are processed with no direct connection to professional or commercial activity. Property rights have been iteratively refined to safeguard personal data privacy and control (Textbox 1). The patient determines who has access to the Genia PSS as a member, the level of interaction with the system, and what patient-controlled real-world data to share. Especially for younger pediatric patients, parents serve as proxies by documenting daily observations about their child [28].

Textbox 1. Seven property rights for patient world arena users of the Genia patient support system.

Access: Two defined user roles have access to the Genia patient support system (PSS) facility: *Patient* (individuals with cystic fibrosis) and *Member* (the innermost network of trusted persons as defined by the patient).

Contribution: Patient and Member users have the right to contribute with data to the patient account they are linked to.

Extraction: Patient and Member users have the right to extract data. Extraction is supported by functionalities in the Genia PSS. The user first selects the data items to extract and generates a report that can be shared. Extractions are logged.

Removal: Patient and Member users have the right to remove data from the patient account they are linked to.

Management/participation: The right to regulate general use patterns and make changes to the Genia PSS is retained by the service provider. Patient and Member users can personalize their individual patient account and activate available feature bundles.

Exclusion: Patient users determine who has the right to access, contribute, extract, and remove data, and how these rights may be transferred to Member users. Certain exclusion rights have been transferred to the service provider under the user agreement. For example, to maintain the integrity of the user's private domain, the service provider has decided to limit the facility to private use, excluding professional or commercial use. This ensures the applicability of legislation that protects the freedom of the domestic domain, medical device regulations, and consumer protection regulations.

Alienation: Alienation rights to the Genia PSS facility (application and servers) have been retained by the service provider, that is, the right to sell or lease management/participation rights. The Patient user has alienation rights to the data, that is, to sell or lease extraction and exclusion rights and how those rights may be transferred.

Clinical Microsystem Arena

The *clinical microsystem arena* comprises the professional sphere at health care clinics. Clinical microsystems are the small, functional, front-line units that provide direct clinical services to patients, that is, the place where health care professionals and patients meet. Health care professionals manage patient care by following clinical practice guidelines that are agreed upon by the professions involved and accepted by their patients. In Sweden, patients meet with their care teams for quarterly check-ups at one of the country's four CF centers, or, for patients who live outside commuting distance, at one of over 60 care centers that collaborate in "shared care" with a designated CF center. All annual check-ups are carried out at CF centers, which represent different clinical microsystem arenas. They are essential building blocks of the larger health system that supports CF patients and the university hospital organizations they are part of.

Collection and Use of Clinical Microsystem Artifacts

Actors of the clinical microsystem arena collect and store patient data in the various EHR systems. If granted access by patients, they may also import patient-controlled real-world data from the Genia PSS. The data are accessed through an export interface in the Genia PSS and imported into a specific module of the EHR, which will here be described as the Genia decision support system (Genia DSS). Once imported into the Genia DSS, the data can be processed according to the rules of the clinical microsystem. At each of the four national CF centers, functional features for processing patient-controlled real-world data that support daily clinical practice and care at home were developed and tested iteratively through small-scale design cycles with health care professionals and patient and family representatives.

Sharing of Clinical Microsystem Artifacts With Other Arenas

The Genia DSS is interoperable with the Genia PSS, as well as the national quality registry through a health information exchange service. To share patient data stored in the EHR system with the patient or the national quality registry, data export areas have been implemented in the Genia DSS, similar to those in the Genia PSS. The export area of the Genia DSS hosts data that have been approved and released for sharing and has an interface that can be accessed by the Genia PSS. For data that are shared with the national quality registry, deidentification of sensitive personal data (eg, through pseudonymization) may be required prior to making data accessible to quality registry users, so that individuals cannot be identified.

Property Rights

The clinical microsystem arena practices under a number of laws, regulations, and guidelines (eg, health information technology and patient safety legislation, and clinical practice guidelines). Therefore, the rules that govern health care professionals' use of patient-controlled real-world data must be aligned with these. Property rights that have been applied in rule setting in the Genia DSS for protecting patients' privacy and control are described in Textbox 2.

Textbox 2. Seven property rights for clinical microsystem arena users of the Genia decision support system.

Access: The Genia decision support system (DSS) can primarily be accessed by health care professional (HCP) users (certified health care professionals). Patients have no user accounts in the Genia DSS, but they have access to designated computer terminals at the cystic fibrosis (CF) centers, with a data input and dashboard module. Also, HCPs can present patient facing dashboards to them during clinical visits.

Contribution: HCP users have the right to contribute data to their patients' records in the Genia DSS. Patients can contribute data through patient-facing input modules at CF center computer terminals.

Extraction: HCP users have the right to extract data, which are constrained by rules in the clinical microsystem, patient consent, as well as guidelines regarding data privacy and patient safety. Extraction is done through standardized report formats. Extractions are logged for auditing purposes.

Removal: Removal rights are restricted in ways similar to accounting systems. Data can be modified and deleted from the user's view in the Genia DSS. For data to be irreversibly deleted from the system, special legislation might apply, such as a Swedish regulation that requires a court decision initiated by the patient. Removal activities are logged for auditing purposes.

Management/participation: The right to regulate general use patterns and make changes to the Genia DSS is retained by the service providers. These rights are influenced by design recommendations and the given health care context, as well as potential constraints by other systems in use.

Exclusion: Generally, the head of the clinic has exclusion rights.

Alienation: Alienation rights to the Genia DSS facility (application and servers) have been retained by the service providers, that is, the right to sell or lease management/participation rights. The controller of data (in general, the clinical department) has alienation rights to patient data artifacts, that is, to sell or lease extraction and exclusion rights and how those rights may be transferred.

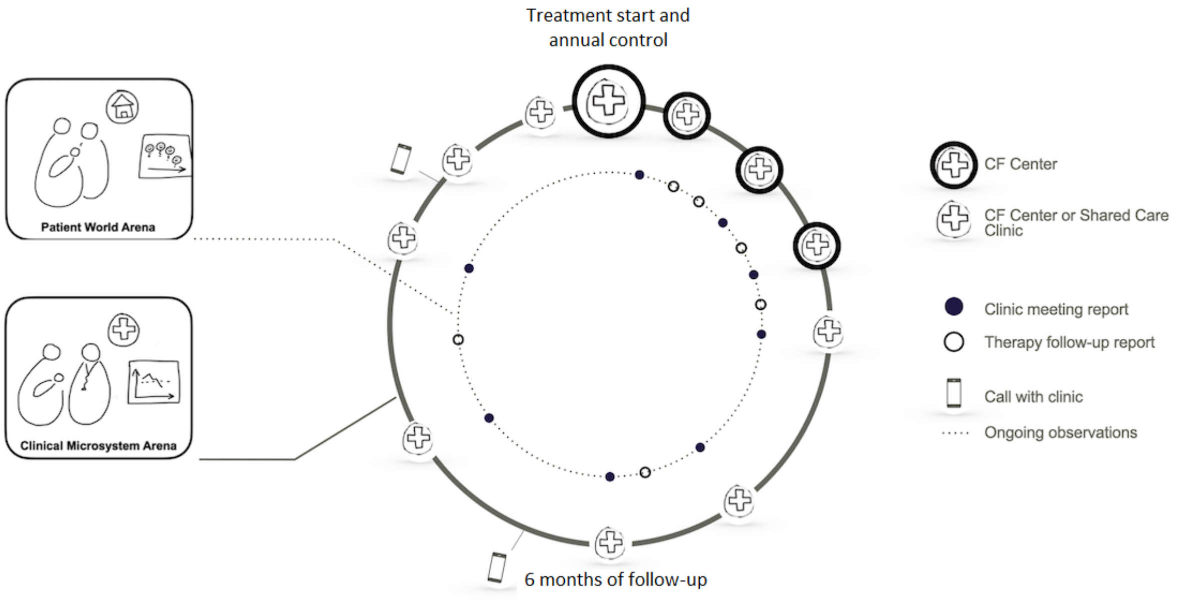
Round Table Arena

The *round table arena* can be defined as a multistakeholder collaborative learning network that works toward common goals. The members of the round table arena may represent individuals (eg, a patient) or institutions (eg, a hospital, authority, or industry actor). A round table arena that includes stakeholders with varying expertise and experience provides distinct advantages, not least of which is the ability to gather and share lessons in real time, contributing to the generation of intellectual and social capital. The Sweden CF Coalition and its Coordination Group are examples of round table arenas that require collective action, that is, the efforts of more than two individuals, to reach their purpose. To implement a system for supporting the orderly introduction and follow-up of lumacaftor/ivacaftor, good coordination is required with national health authorities and councils involved in decisions regarding the introduction and subsidy of the new therapy.

Collection and Use of Round Table Artifacts

With data from the national quality registry as a shared resource, as well as the personal expertise and experiences shared by its members, the Sweden CF Coalition aims to set up crucial conversations among stakeholders, so they can learn about new ways to generate optimal health and high-value care. To implement a national system for the orderly introduction and follow-up of lumacaftor/ivacaftor, changes to existing facilities that hold patient-controlled real-world data were necessary. This involved negotiations on what core data sets could be extracted from the Patient World and Clinical Microsystem arenas. New modules for lumacaftor/ivacaftor follow-up visits with indicators for follow-up and evaluation of the therapy were implemented in the Genia PSS, Genia DSS, and national patient registry. The health information exchange service was modified to enable ethically and legally viable sharing of data across arenas. These modifications enabled the provision of a shared resource to support clinical decision making, quality improvement, administration, and research. The system was implemented in the summer of 2018 upon national approval of market access to lumacaftor/ivacaftor (Figure 2).

Figure 2. First-year customer journey with lumacaftor/ivacaftor therapy (an idealized flow with a new therapy linking the patient world and clinical microsystem arenas).



Sharing of Round Table Artifacts With Other Arenas

Actors in the Sweden CF Coalition share their collaboratively generated ideas through analyses, reports, and clinical guidelines. These artifacts are made available to selected stakeholders through creative commons or fee-for-service agreements in accordance with industry standards and national ethical platforms, or generally made available to the public. The information and knowledge that are produced by the round table arena can be described as *common-pool resources*, which means that they are controlled by stakeholders within the Sweden CF Coalition.

Property Rights

In contrast to the patient world arena and the clinical microsystem arena, where conversations about health concern identified individuals, actors in the round table arena have access to data representing deidentified individuals or a nonidentified collective of individuals [36]. Based on different round table arena actors' roles and affiliations to other settings, the information they contribute to and extract from the facilities they have access to will be governed by different property rights. The data made publicly available by the Sweden CF Coalition are from the national patient registry for CF, which serves as an *open-access common-pool resource* where no one has the legal right to exclude anyone from using the resource, as long as it is in conformance with general legal requirements.

Discussion

Principal Findings

This study used an action research approach to propose a solution for enabling patients to effectively share patient-controlled real-world data in CF care. The case that was presented is an example of coproduction in and beyond health care services, where patients contribute to the orderly introduction and follow-up of new therapies, while retaining control of their data. We identified the following three types of action arenas where patient-controlled real-world data are collected, used, and shared: patient world arena, which comprises the private sphere of patients and their families; clinical microsystem arena, which comprises the professional sphere at health care clinics; and round table arena, which comprises stakeholders from multiple spheres with complementary expertise. The relationships and information exchange between these arenas were illustrated in a conceptual model, which we suggest can be used as a tool to analyze current chronic care systems and as a guide for the development of solutions for patient-controlled real-world data sharing.

Comparison With Prior Work

In the following sections, we discuss our results in relation to prior work. We discuss the conceptualization of the three arenas in our model, the sharing of patient-controlled real-world data, and, finally, the applicability of the model in other chronic care settings.

The Three Arenas

The patient world arena is a central building block of the conceptual model presented in this study. In line with well-known concepts, such as *patient-centered care* [4], it emphasizes the patient's experience and expertise on living with a chronic condition. Our model acknowledges that patients may involve actors outside the health care organization in their self-care. Thus, the patient world arena describes the patient's domestic domain where laws and regulations of the health care setting and general privacy directives do not apply [37]. While each patient world arena has its own (likely implicit) social rules and norms for how data are captured, used, and shared, our model identifies some general property rights that may be applied.

The concept of the clinical microsystem, in contrast to the patient world, is well established. Nelson et al [38] described the clinical microsystem as "the local milieu in which patients, providers, support staff, information, and processes converge for the purpose of providing care to meet health needs" [38]. It can be seen as an arena for coproduction between the patient and his or her self-determined network and the formal care team, aimed at promoting, improving, or maintaining individual health. As Ostrom notes [7], "technologies in use must generate a complimentary production possibility frontier […] rather than merely a substitutive one. […] When coproductive inputs are legally owned by diverse entities and complements, synergy can occur." Thus, while the clinical microsystem may be embedded in the health care organization, the patient represents an own entity when appearing as an actor and guest in the clinical microsystem arena. Here, the specification of property rights helps to clarify the ownership and control of the different inputs to the system, where patient-controlled real-world data that are produced in the health care setting or at home are controlled by the patient, but can be shared with the care team for use in the patient's individual care.

The concept of the round table arena builds on the previous work by Lindblad (SL) et al [36,39] on registry-based learning networks, as well as previous research on collaborative chronic care networks [40-42]. It differs from the other arenas in a number of aspects. First, while the patient world and clinical microsystem arenas deal with coproducing strategies to maximize individual health, the round table arena deals with coproducing strategies to maximize population health. The Sweden CF Coalition that was analyzed in this study can be described as a double-loop learning system [43] that engages a multistakeholder group in following up and discussing the adequacy of new therapy practices and related guidelines, based on an aggregation of data from the patient world and clinical microsystem arenas. It deals with quality improvement on a national level. Second, while the patient world and clinical microsystem arenas are found within the boundaries of the domestic domain and health care organization, respectively, the round table arena crosses organizational boundaries. Third, while the purpose and stakeholders of the patient world and clinical microsystem arenas are well known to the patient, the round table arena may be formed for a variety of purposes and in a variety of settings. The patient who contributes with data may not know the stakeholders and may not gain any direct benefit that contributes to his or her individual care, at least not immediately. Thus, communication of purposes and agreements on the terms of patient-controlled real-world data management may be all the more important.

Sharing of Patient-Controlled Real-World Data Between Arenas

In this action research case study, we identified several different facilities that were used by stakeholders to share and access patient-controlled real-world data (the Genia PSS, Genia DSS, and quality registry). Sweden has long experience with clinical quality registries that allow patient-valued health outcomes and research to be coproduced between different stakeholders and provide open access to quality registry data for quality improvement and benchmarking purposes [36,44,45]. Previous research has shown that the majority of people are willing to share their personal health data, particularly for research purposes [46].

The proliferation of pervasive computing technologies, such as the "internet of things," will greatly increase data collection from the personal and private spheres, which may be conducive to social dilemmas such as disempowerment and enclosure. This all the more emphasizes the need to protect the integrity of individuals and their private life, which is key to support the success of such technologies [47]. Legal requirements like the GDPR have strengthened individuals' rights to the protection of personal data, which, in our opinion, makes patients the only ones who can make sustainable decisions about the sharing of patient-controlled real-world data. Consequently, patients should be in the driver's seat when these data are utilized to develop, distribute, and evaluate new medical technologies, care processes, and systems. By applying the seven property rights suggested by Ostrom and Hess [25] in this study, rules were specified to protect individuals' privacy in line with moral and legal requirements. For example, the access, contribution, extraction, removal, and alienation rights to patient-controlled real-world data in the Genia PSS were exclusively specified for the patient and his or her self-determined network. The service providers' rights were mainly limited to aspects that concern the facility, and its functionalities and use patterns, but no rights to patient-controlled real-world data were specified, which prevents uncontrolled data analytics.

Applicability to Other Chronic Care Settings

Although the model presented in this study is based on the context of CF care, we believe that it is applicable to other chronic care settings that share similarities with CF care in terms of digital technology use to support coproduction. We selected CF care as the context for this case study because of its strong dependence on coproduction between patients, their families, and health care [20,21]. This type of coproduction is not unique to CF care and we argue that, in light of the demographic shift and technological development that we described in the introduction section of this paper, the transition to co-care in chronic care management [8] is a necessary paradigm shift in the organization and management of health care. Apart from CF care, we have observed other examples of chronic care settings where existing technological, legal, and social structures can be analyzed on the basis of the conceptual model we presented in this study. For example, the round table arena concept has previously been described in the context of rheumatoid arthritis [36]. Similarly, the idea of collaborative chronic care networks stems from research on inflammatory bowel disease [40-42]. Further, the Genia PSS has already spread to other chronic conditions in Sweden (juvenile idiopathic arthritis [28] and acute intermittent porphyria), which allows us to speculate that our model involving three action arenas may be applicable at least in these settings and likely other chronic care settings where co-care is aspired.

Initially, the model can be used to raise questions like “What action arenas can we identify in the chronic care management of a certain patient population/in a certain setting?” “Who are (potential) members of the different arenas?” “Which norms and rules apply in the interactions within and between arenas?” “What real-world data flows exist between arenas?” “What facilities are in use?” and “What property rights have been specified and implemented to protect patients’ privacy and control?” We believe that the discussions triggered by these types of questions may help to uncover some of the social dilemmas that may result from data use and sharing not controlled by the patient, such as disempowerment and enclosure [16]. Such dilemmas may be addressed by formulating and implementing appropriate property rights. In summary, we believe that this model may find application as a tool for the analysis and comparison of chronic care systems in different patient groups and settings, as well as a guide for the development of appropriate technical and legal structures to support coproduction between different arenas.

Limitations

This study is not without limitations. The model involving three action arenas that was presented resulted from an iterative cycle of development and analysis based on a single case. Because our intention was to present a model that is also applicable in other chronic care settings, we chose a relatively high level of abstraction, compromising the level of detail. We see our model as a general guide rather than a detailed manual on how to manage the sharing of patient-controlled data in chronic care settings. We acknowledge that AH had a central role in both the development and analysis. To maintain neutrality from preconceived ideas or personal interests, the analysis and results have been discussed among all the authors, guided by theory. To further develop the model and test its transferability to other settings, pilot studies have been initiated in the United States (Clinical Trial ID: NCT03910881) and Argentina, and we plan to conduct an international multiple case study.

Conclusions

To date, little academic research has been devoted to the study of how patient-controlled real-world data can be employed in the introduction and systematic evaluation of novel health technologies. Health care professionals, health care funders, researchers, authorities, and the industry are all dependent on the collection and analysis of real-world data reported by the individual patient whose personal data privacy and control need to be protected. To proactively manage social dilemmas related to shared information and knowledge resources, we suggest the specification of property rights that enable patients to determine how their personal health data are managed and shared in other settings. Our model, developed as a knowledge commons, comprises technical and legal structures that protect the integrity and self-determination of patients in such data sharing to improve individual and population health. Further research is needed to adapt and transfer our proposed knowledge commons model to improve health in other chronic conditions as well.

Acknowledgments

This study is funded by Merck Sharp & Dohme Corp through a strategic partnership between Karolinska Institutet and the company, encouraging collaboration focusing on patient-centered research. AH and CW were also funded by the Swedish Research Council and the Swedish Research Council for Health, Working Life and Welfare (grant number 2018-01472 and 2017-01451), and MB was supported by the Swedish Research Council for Health, Working Life and Welfare (grant number 2014-04238). The authors extend their gratitude to the Sweden CF Coalition, the cystic fibrosis community, Hector Gutierrez, Gabriela Oates, Manolis Nymark, and Eugene Nelson for their valuable support, as well as the reviewers for their valuable comments on several versions of the manuscript.

Authors' Contributions

AH and CW drafted the final manuscript, with contributions from SL and MB. All authors contributed to the study design and analysis, and approved the final version of the manuscript.

Conflicts of Interest

AH is a father of children with cystic fibrosis and an eHealth entrepreneur engaged in developing the Genia patient support system and health information exchanges. All other authors declare no conflicts.

References

1. Wammes JJG, van der Wees PJ, Tanke MAC, Westert GP, Jeurissen PPT. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open 2018 Sep 08;8(9):e023113 [FREE Full text] [doi: 10.1136/bmjopen-2018-023113] [Medline: 30196269]
2. Greenhalgh T. Patient and public involvement in chronic illness: beyond the expert patient. BMJ 2009 Feb 17;338:b49. [doi: 10.1136/bmj.b49] [Medline: 19223339]
3. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]

4. Hudon C, Fortin M, Haggerty J, Loignon C, Lambert M, Poitras M. Patient-centered care in chronic disease management: a thematic analysis of the literature in family medicine. Patient Educ Couns 2012 Aug;88(2):170-176. [doi: 10.1016/j.pec.2012.01.009] [Medline: 22360841]
5. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
6. Batalden M, Batalden P, Margolis P, Seid M, Armstrong G, Opipari-Arrigan L, et al. Coproduction of healthcare service. BMJ Qual Saf 2016 Jul;25(7):509-517 [FREE Full text] [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]
7. Ostrom E. Crossing the great divide: Coproduction, synergy, and development. World Development 1996 Jun;24(6):1073-1087 [FREE Full text] [doi: 10.1016/0305-750x(96)00023-x]
8. von Thiele Schwarz U. Co-care: Producing better health outcome through interactions between patients, care providers and information and communication technology. Health Serv Manage Res 2016 Apr 04;29(1-2):10-15. [doi: 10.1177/0951484816637746]
9. Christensen CM, Bohmer R, Kenagy J. Will disruptive innovations cure health care? Harv Bus Rev 2000;78(5):102-12, 199. [Medline: 11143147]
10. Hoy MB. Personal Activity Trackers and the Quantified Self. Med Ref Serv Q 2016;35(1):94-100. [doi: 10.1080/02763869.2016.1117300] [Medline: 26794199]
11. European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. 2016 May 4. URL: http://data.europa.eu/eli/reg/2016/679/oj [accessed 2021-01-05]
12. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. The Learning Healthcare System: Workshop Summary. The National Academies Collection: Reports funded by National Institutes of Health 2007 Jan 01:1-354. [doi: 10.17226/11903] [Medline: 21452449]
13. Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff (Millwood) 2014 Jul;33(7):1163-1170 [FREE Full text] [doi: 10.1377/hlthaff.2014.0053] [Medline: 25006142]
14. Hess C, Ostrom E, editors. Understanding Knowledge as a Commons: From Theory to Practice. Cambridge, Massachusetts: MIT Press; 2006.
15. Hess C, Ostrom E. Introduction: An Overview of the Knowledge Commons. In: Hess C, Ostrom E, editors. Understanding Knowledge as a Commons: From Theory to Practice. Cambridge, Massachusetts: MIT Press; 2007:3-26.
16. Purtova N. Health Data for Common Good: Defining the Boundaries and Social Dilemmas of Data Commons. In: Adams S, Purtova N, Leenes R, editors. Under Observation: The Interplay Between eHealth and Surveillance. Law, Governance and Technology Series, vol 35. Cham, Switzerland: Springer; 2017:177-213.
17. Bollier D. The Growth of the Commons Paradigm. In: Hess C, Ostrom E, editors. Understanding Knowledge as a Commons: From Theory to Practice. Cambridge, Massachusetts: MIT Press; 2007:27-40.
18. Parkin P. Action Research as a Strategy for Implementing Change. In: Managing Change in Healthcare: Using Action Research. Thousand Oaks, California: SAGE Publications; 2009:14-32.
19. Meyer J. Qualitative research in health care. Using qualitative methods in health related action research. BMJ 2000 Jan 15;320(7228):178-181 [FREE Full text] [doi: 10.1136/bmj.320.7228.178] [Medline: 10634744]
20. Sabadosa KA, Batalden PB. The interdependent roles of patients, families and professionals in cystic fibrosis: a system for the coproduction of healthcare and its improvement. BMJ Qual Saf 2014 Apr;23 Suppl 1:i90-i94. [doi: 10.1136/bmjqs-2013-002782] [Medline: 24608555]
21. Elborn JS. Cystic fibrosis. The Lancet 2016 Nov 19;388(10059):2519-2531. [doi: 10.1016/S0140-6736(16)00576-6] [Medline: 27140670]
22. Smyth AR, Bell SC, Bojcin S, Bryon M, Duff A, Flume P, European Cystic Fibrosis Society. European Cystic Fibrosis Society Standards of Care: Best Practice guidelines. J Cyst Fibros 2014 May;13 Suppl 1:S23-S42 [FREE Full text] [doi: 10.1016/j.jcf.2014.03.010] [Medline: 24856775]
23. McColley SA, Konstan MW, Ramsey BW, Stuart Elborn J, Boyle MP, Wainwright CE, et al. Lumacaftor/Ivacaftor reduces pulmonary exacerbations in patients irrespective of initial changes in FEV1. Journal of Cystic Fibrosis 2019 Jan;18(1):94-101 [FREE Full text] [doi: 10.1016/j.jcf.2018.07.011] [Medline: 30146268]
24. de Monestrol I, Lindblad A. CF-registret Årsrapport 2018 (CF register Annual Report 2018). CF-registret: ett svenskt kvalitetsregister (The CF registry: a Swedish quality register). 2019 Sep. URL: https://cf-registret.se/onewebmedia/%C3%85rsrapport%202108%20webversion%2020190913.pdf [accessed 2021-01-05]
25. Ostrom E, Hess C. A Framework for Analyzing the Knowledge Commons. In: Hess C, Ostrom E, editors. Understanding Knowledge as a Commons: From Theory to Practice. Cambridge, Massachusetts: MIT Press; 2007:41-81.
26. Hess C, Ostrom E. Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource. Law and Contemporary Problems 2003;66(1/2):111-145 [FREE Full text]

27. Longacre M, Grande S, Hager A, Montan M, Bergquist RP, Martensson M, et al. Clinical Adoption of mHealth Technology to Support Pediatric Cystic Fibrosis Care in Sweden: Qualitative Case Study. JMIR Pediatr Parent 2018 Dec 10;1(2):e11080 [FREE Full text] [doi: 10.2196/11080] [Medline: 31518297]
28. Grande SW, Longacre MR, Palmblad K, Montan MV, Berquist RP, Hager A, et al. Empowering Young People Living With Juvenile Idiopathic Arthritis to Better Communicate With Families and Care Teams: Content Analysis of Semistructured Interviews. JMIR Mhealth Uhealth 2019 Feb 22;7(2):e10401 [FREE Full text] [doi: 10.2196/10401] [Medline: 30794202]
29. Sundeson K. See how Genia works, meet Victor. YouTube. 2016 Feb 21. URL: https://youtu.be/rJQu2YTHr6s [accessed 2021-01-05]
30. Sundeson K. Collaboration in care: Design of a mobile health platform to support care management of cystic fibrosis. Aalto University. 2014. URL: https://aaltodoc.aalto.fi/bitstream/handle/123456789/13047/master_Sundeson_Kajsa_2014.pdf?sequence=1&isAllowed=y [accessed 2021-01-05]
31. de Monestrol I, Ericson P, Hjelte L, Mostrom J, Hager A, Lindblad S. Co-creating the Swedish example for orderly introduction and follow-up of new therapies. In: Pediatric Pulmonology Vol. 53. 2018 Presented at: The 32nd Annual North American Cystic Fibrosis Conference; October 18-20, 2018; Denver, Colorado p. S148-S456.
32. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
33. Dental and Pharmaceutical Benefits Agency. Orkambi ingår i läkemedelsförmånerna [Orkambi included in drug subsidy]. Dental and Pharmaceutical Benefits Agency, TLV–Decisions. 2018. URL: https://www.tlv.se/download/18.500ea4181641067957a31b99/1529412747506/bes180614_orkambi.pdf [accessed 2021-01-05]
34. Managed introduction – this is how it works. Janusinfo. 2018 Apr 17. URL: https://janusinfo.se/nationelltinforandeavlakemedel/managedintroductionthisishowitworks.4.6b32b8ec162bd970d6bb885.html [accessed 2021-01-05]
35. Kleinman A. Illness Meanings and Illness Behaviour. In: McHugh S, Vallis TM, editors. Illness Behavior. Boston, Massachusetts: Springer; 1986:149-160.
36. Lindblad S, Ernestam S, Van Citters AD, Lind C, Morgan TS, Nelson EC. Creating a culture of health: evolving healthcare systems and patient engagement. QJM 2017 Mar 01;110(3):125-129. [doi: 10.1093/qjmed/hcw188] [Medline: 27803364]
37. Agerberg M. Journal på Nätet ska bli del i hälsopaket [Online EHR will become part of the health package]. Läkartidningen 2012 Feb 7;109(6):266-271 [FREE Full text]
38. Nelson EC, Batalden PB, Huber TP, Mohr JJ, Godfrey MM, Headrick LA, et al. Microsystems in Health Care: Part 1. Learning from High-Performing Front-Line Clinical Units. The Joint Commission Journal on Quality Improvement 2002 Sep;28(9):472-493. [doi: 10.1016/s1070-3241(02)28051-7] [Medline: 12216343]
39. Ovretveit J, Keller C, Hvitfeldt Forsberg H, Essén A, Lindblad S, Brommels M. Continuous innovation: developing and using a clinical database with new technology for patient-centred care--the case of the Swedish quality register for arthritis. Int J Qual Health Care 2013 Apr;25(2):118-124. [doi: 10.1093/intqhc/mzt002] [Medline: 23360809]
40. Margolis PA, Peterson LE, Seid M. Collaborative Chronic Care Networks (C3Ns) to transform chronic illness care. Pediatrics 2013 Jun;131 Suppl 4:S219-S223 [FREE Full text] [doi: 10.1542/peds.2012-3786J] [Medline: 23729764]
41. Seid M, Dellal G, Peterson LE, Provost L, Gloor PA, Fore DL, et al. Co-Designing a Collaborative Chronic Care Network (C3N) for Inflammatory Bowel Disease: Development of Methods. JMIR Hum Factors 2018 Feb 22;5(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.8083] [Medline: 29472173]
42. McLinden D, Myers S, Seid M, Busch M, Davis D, Murphy J. The Learning Exchange, a Community Knowledge Commons for Learning Networks: Qualitative Evaluation to Test Acceptability, Feasibility, and Utility. JMIR Form Res 2019 Mar 14;3(1):e9858 [FREE Full text] [doi: 10.2196/formative.9858] [Medline: 30869643]
43. Argyris C. Double loop learning in organizations. Harvard Business Review. 1977. URL: https://hbr.org/1977/09/double-loop-learning-in-organizations [accessed 2021-01-05]
44. Hvitfeldt H, Carli C, Nelson EC, Mortenson DM, Ruppert BA, Lindblad S. Feed forward systems for patient participation and provider support: adoption results from the original US context to Sweden and beyond. Qual Manag Health Care 2009;18(4):247-256. [doi: 10.1097/QMH.0b013e3181bee32e] [Medline: 19851232]
45. Nelson EC, Dixon-Woods M, Batalden PB, Homa K, Van Citters AD, Morgan TS, et al. Patient focused registries can improve health, care, and science. BMJ 2016 Jul 01;354:i3319 [FREE Full text] [doi: 10.1136/bmj.i3319] [Medline: 27370543]
46. Karampela M, Ouhbi S, Isomursu M. Connected Health User Willingness to Share Personal Health Data: Questionnaire Study. J Med Internet Res 2019 Nov 27;21(11):e14537 [FREE Full text] [doi: 10.2196/14537] [Medline: 31774410]
47. Article Data Protection Working Party. Opinion 8/2014 on the on Recent Developments on the Internet of Things. European Commission. 2014. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf [accessed 2021-01-05]

Abbreviations

CF: cystic fibrosis

DSS: decision support system
EHR: electronic health record
GDPR: General Data Protection Regulation
PSS: patient support system

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by I Mircheva, S Sarbadhikari, J Jones; comments to author 10.03.20; revised version received 15.06.20; accepted 17.12.20; published 19.01.21.

Please cite as:
Hager A, Lindblad S, Brommels M, Salomonsson S, Wannheden C
Sharing Patient-Controlled Real-World Data Through the Application of the Theory of Commons: Action Research Case Study
J Med Internet Res 2021;23(1):e16842
URL: http://www.jmir.org/2021/1/e16842/
doi:10.2196/16842
PMID:33464212

Original Paper

Innovation in Pain Rehabilitation Using Co-Design Methods During the Development of a Relapse Prevention Intervention: Case Study

Stefan Elbers[1,2], MSc; Christa van Gessel[3], MSc; Reint Jan Renes[4], PhD; Remko van der Lugt[3], PhD; Harriët Wittink[1], PhD; Sander Hermsen[1,5], PhD

[1]Research Group Lifestyle and Health, Utrecht University of Applied Sciences, Utrecht, Netherlands

[2]Department of Rehabilitation Medicine, Faculty of Health, Life Sciences and Medicine, Maastricht University, Maastricht, Netherlands

[3]Co-design Research Group, Utrecht University of Applied Sciences, Utrecht, Netherlands

[4]Research Group Psychology of Sustainable Cities, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[5]OnePlanet Research Center, imec The Netherlands, Wageningen, Netherlands

Corresponding Author:
Stefan Elbers, MSc
Research Group Lifestyle and Health
Utrecht University of Applied Sciences
Heidelberglaan 7
Utrecht, 3508 AA
Netherlands
Phone: 31 88 481 78 69
Email: stefan.elbers@hu.nl

Abstract

Background: Many intervention development projects fail to bridge the gap from basic research to clinical practice. Instead of theory-based approaches to intervention development, co-design prioritizes the end users' perspective as well as continuous collaboration between stakeholders, designers, and researchers throughout the project. This alternative approach to the development of interventions is expected to promote the adaptation to existing treatment activities and to be responsive to the requirements of end users.

Objective: The first objective was to provide an overview of all activities that were employed during the course of a research project to develop a relapse prevention intervention for interdisciplinary pain treatment programs. The second objective was to examine how co-design may contribute to stakeholder involvement, generation of relevant insights and ideas, and incorporation of stakeholder input into the intervention design.

Methods: We performed an embedded single case study and used the double diamond model to describe the process of intervention development. Using all available data sources, we also performed deductive content analysis to reflect on this process.

Results: By critically reviewing the value and function of a co-design project with respect to idea generation, stakeholder involvement, and incorporation of stakeholder input into the intervention design, we demonstrated how co-design shaped the transition from ideas, via concepts, to a prototype for a relapse prevention intervention.

Conclusions: Structural use of co-design throughout the project resulted in many different participating stakeholders and stimulating design activities. As a consequence, the majority of the components of the final prototype can be traced back to the information that stakeholders provided during the project. Although this illustrates how co-design facilitates the integration of contextual information into the intervention design, further experimental testing is required to evaluate to what extent this approach ultimately leads to improved usability as well as patient outcomes in the context of clinical practice.

(J Med Internet Res 2021;23(1):e18462) doi:10.2196/18462

KEYWORDS

co-design; participatory design; chronic pain; intervention development; rehabilitation; behavior change; relapse; prevention

Introduction

Only a fraction of intervention development projects is able to bridge the translational gap from scientific research to clinical practice [1-4]. An important factor for this limited uptake may be that contextual factors, such as stakeholder acceptability or implementation in existing practices, receive little attention during earlier development stages [5]. For example, many intervention development guidelines emphasize the formulation of an underlying theoretical construct and subsequent experimental testing to validate each assumed causal step, but only address implementation and feasibility after the intervention development phase has been completed [6-8]. Consequently, theoretically sound interventions may be discarded due to insufficient attention to crucial translational factors such as low perceived utility by patients or health care providers (HCPs), inconvenient navigation, or a discrepancy between the intervention requirements and patients' preferences [9-14].

An opportunity to increase the emphasis on these factors is to incorporate co-design methods. Co-design not only is characterized by an incremental knowledge over multiple development cycles [15] but also specifically emphasizes empathizing with each stakeholder, integrating conflicting requirements, and quickly transitioning ideas to testable prototypes. Co-design differs from other design methodologies in that it involves a range of tools and exercises to optimize collaboration between professional designers and people who are not trained in the design process, such as patients and therapists [16]. Done rightly, co-design brings together different views, input, and competencies of people with a variety of perspectives to address a specific problem [17,18]. As a result, this approach should increase the acceptability and integration of the intervention in existing clinical practice by accommodating relevant contextual factors that have been identified by stakeholders in the development process.

Although co-design is increasingly adopted in the development of health care interventions (eg, [18-23]), prior studies have indicated that effective co-design is not without its challenges. For example, the process of engaging all stakeholders can be time-consuming and intensive. This can be particularly difficult in the context of health care because HCPs generally have a high workload [20], and participating patients often do not directly benefit from the development projects, which could negatively influence their motivation and engagement. Moreover, patients, policy makers, and HCPs can experience conflicting interests during intervention development projects, because the assumed best possible care is generally limited by finite resources or specific treatment guidelines within a particular health care environment [24]. Factors such as these could endanger the main principles of co-design and should be further examined in the context of health care [18,19].

Co-Design in the Context of Chronic Pain

In the present project, called the SOLACE project (grant number: SIA RAAK 2014-01-23), we developed a relapse prevention intervention for patients with chronic musculoskeletal pain who participate in an interdisciplinary, multimodal pain treatment program. The primary reason for adopting a co-design approach was that, despite high prevalence rates of relapse after successful treatment, there is a paucity of available research to explain relapse for this particular population [25,26]. In these situations, a design-based approach may be particularly appropriate, because it allows for new insights to be recursively fed into future development cycles, thereby gradually increasing the knowledge base over time [15,27]. Because patients with chronic pain often experience distrust from their personal and medical environment [28], co-design may also prove effective in empowering patients to participate in the development process and to actively share their opinions and ideas [23,29].

Objectives

To increase understanding of how co-design can be successfully applied in the development of interventions in the health care domain, more examples of good practice are needed [13,18,30]. Therefore, our research question was to what extent co-design practices facilitate the translation of meaningful stakeholder experiences into the design of a health care intervention. Our first aim was to provide a detailed overview of all co-design activities that were employed during the course of an example project. Our second aim was to reflect on this overview and examine how co-design may contribute to stakeholder involvement, generation of relevant insights and ideas, and incorporation of stakeholder input into the intervention design. We hypothesized that co-design activities would facilitate the generation of relevant experiences and insights from stakeholders and stimulate their active participation during this project. Consequently, we expected that this would yield prototypes that were aligned with clinical practice and would resonate with end users.

Methods

Design

We performed an embedded single case study [31], in which we analyzed and evaluated all co-design activities that were related to the development process of the SOLACE project. Throughout the study, the researchers followed a participatory action research (PAR) approach, which is characterized by active collaboration with the people of interest, rather than only researching them. PAR also emphasizes respectful cooperation between stakeholders and researchers including collective decision making and a bidirectional transfer of knowledge over multiple iterative development cycles [32-34], which is in accordance with co-design methods [35]. During each cycle, insight is acquired through action (eg, by letting patients interact with preliminary prototypes), empowerment of stakeholders (eg, by patient involvement in co-design sessions), and subsequent reflection [32].

Participants

The SOLACE project consortium consisted of 2 interdisciplinary multimodal pain treatment centers, the Royal Dutch Society for Physical Therapy, The Dutch National Pain Patient Advocacy Organisation, and 4 research groups with a respective interest in chronic pain treatment (2 groups), co-design, and behavior change. All consortium partners assisted with the recruitment of participants when this was required at specific co-design activities, including patients and their spouses, HCPs, designers, researchers, and students. The core team was composed of 3 researchers, each from a different research group. This team was responsible for the planning and preparation of the co-design activities. To monitor overall progress, a steering committee was formed, which included representatives of all consortium partners. Ethical approval for this study was granted by the local ethics committee (Medical Research Ethics Committee Atrium 15-N-120).

Materials

Co-Design Methods

In interviews and co-creation sessions, the core team adopted various co-design methods, including generative techniques, contextual interviews, system mapping, and prototyping. These methods were adopted to facilitate stakeholder participation during key moments in the design process: generative techniques to elicit tacit knowledge and latent needs, contextual interviews to increase empathy, system mapping to develop a comprehensive overview of the acquired data, and prototyping to make ideas tangible and possible to experience. Co-creation sessions included multiple co-design methods and were specifically employed to empower a variety of stakeholders to participate in the design process.

Semistructured Contextual Interviews

At various time points in the project, we interviewed patients and HCPs. The interviews were performed by 2 researchers and were conducted in the everyday context of the HCPs (treatment facility) and patients (at home). To activate prior knowledge and experiences, all participants received “sensitizers”—assignments that stimulated thinking about relevant topics—before the interview (see page 4 in Multimedia Appendix 1) [36]. During the interview, the primary interviewer explored participants’ beliefs, needs, and experiences using open questions and various generative techniques. The second interviewer took notes and asked additional questions to ensure that all topics that the research team identified during preparatory sessions were covered. Data were collected by audio recording and note taking. Directly after the interview, both interviewers discussed their impressions and updated their notes.

Generative Techniques

To explore participants’ ideas, needs, and values beyond their first response, various generative techniques were employed during interview and design sessions. These techniques aim to bring up “tacit knowledge” by addressing social, emotional, and functional elements related to a topic of interest [36]. For example, to promote a more personal acquaintance during the interview sessions, participants were asked to introduce themselves by selecting 3 pictures from a deck of cards illustrated with ambiguous images that symbolized their personal values. The core team also used journey mapping during interviews (see page 3 in Multimedia Appendix 1). This technique enabled all attendees to collaboratively construct a graphic visualization or a timeline that illustrated their experiences with interdisciplinary multimodal pain treatment [37,38].

Prototyping and Provotyping

A key element of PAR is to increase insight by reflecting on actual interactions with prototypes. As Step 1.3 on page 5 of Multimedia Appendix 1 illustrates, the interaction with these objects stimulated the individual to envision future possibilities or to visualize concept ideas. The process of prototyping allows participants to actively engage with objects that were based on preliminary outcomes and encompass operationalizations of the concept of interest [39]. Provotyping takes place with objects that are not directly related to the final result but are specifically designed to test a specific hypothesis or elicit a particular response [40].

System Maps and Personas

System mapping is a method for creating a visual representation of interacting variables that facilitates the understanding of complex systems [41]. System maps typically include a framework of interrelated components, as well as clarifying examples of quotes and pictures. These maps can be presented in posters or cards and are useful to share data to stakeholders in an accessible way (see pages 7-9 in Multimedia Appendix 1). Moreover, it provides participants with the opportunity to jointly reflect on the data and update ideas during co-creation sessions.

A specific way to represent the data as a coherent “whole” for usage throughout co-design activities is by creating personas: fictitious archetypes of users, each reflecting a distinct pattern in goals, attitudes, and behaviors based on empirical research among potential users. With personas, it is possible to highlight certain areas of tension or to facilitate discussion of important patient characteristics [42].

Co-Creation Sessions

We used co-creation sessions at key moments during the project to discuss and reflect on the collected data, to generate new ideas, and to make decisions regarding future development directions (see page 10 in Multimedia Appendix 1). A typical session would take 4 hours and involved 10-20 stakeholders. The core team prepared the sessions by formulating desired outcomes and setting up system maps to present the data. During the sessions, 2 designers operated as workshop facilitators and used various assignments (eg, prioritizing ideas for prototype concepts) to work towards the desired outcomes in an open atmosphere where everyone was invited to actively participate. All written session data (eg, posters, drawings, notes) were collected and discussed during core team evaluation meetings directly after the session. To maintain involvement and commitment between the sessions, the core team sent bimonthly newsletters and posted updates on the project websites.

Measurements and Analysis

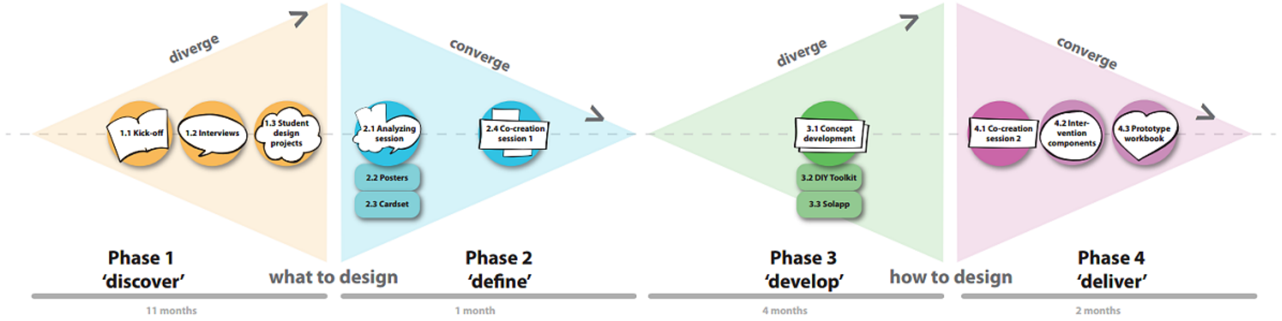
Dataset

The dataset for this case study consisted of 4 different sources. To capture the results of the design methods, researchers documented each design and research activity, using observation notes, pictures, audio files, or video clips. In addition to the session documentation, researchers also organized reflective sessions directly after a co-design activity to summarize the output of co-creation sessions (eg, notes or post-its) into system maps. These maps served both as a descriptive analysis of the data as well as for input during subsequent co-design sessions. The dataset also consisted of minutiae of steering committee meetings and a retrospective project journey. This journey was the result of a reviewing session, where researchers and members of the steering committee chronologically described and discussed critical incidents.

Data Analysis

We used a deductive content approach to identify information within the dataset that relates to our main themes: stakeholder involvement, generation of insights, and incorporation of stakeholder input. We defined stakeholder involvement as the commitment to participate in the development project, to collaborate with other stakeholders during design activities, and to actively participate during these sessions. Generation of insights referred to the extent by which co-design activities resulted in an increased understanding of the problem of interest that could inform subsequent development activities. Incorporation of stakeholder needs was defined as the extent by which prototypes were based on stakeholder perceptions, judgments, and evaluations from co-design activities. Furthermore, we adopted the Double Diamond model to describe the design process along 4 development stages [43]. The Double Diamond model contains 2 sequences of diverging and converging (Figure 1). In diverging phases, choice options and discrepancies are created; in the converging phases, these ideas are refined and considered to make design choices with respect to the prototype. In the first diamond, the “Discover” (diverging) and “Define” (converging) phases relate to acquiring insights on *what* to design. In the second diamond, the “Develop” and “Deliver” relate to further exploring the ideas on *how* to optimally design the final concepts. To illustrate how co-design contributed to the intervention development at each phase, we combined all data sources to provide both a descriptive overview and an in-depth reflection with respect to our main themes. In addition, Multimedia Appendix 1 provides a chronological overview of the development process and includes examples of co-design methods, data segments, and pictures of co-design activities.

Figure 1. Overview of the co-design development process.



Results

Phase 1: “Discover’”

Description

In the “Discover” phase, we aimed to generate a deeper understanding of factors influencing relapse after successful rehabilitation. The primary activities took place over a period of 11 months and consisted of 3 kick-off sessions, 20 stakeholder interviews (12 HCPs, 8 patients), and a student design project. In the first kick-off session, representatives from all consortium partners were present to discuss the project planning and to decide how co-design would be implemented throughout the project. Representatives also participated in co-creation by using their professional and personal experiences to formulate initial ideas on relapse (see page 3 of Multimedia Appendix 1). These activities were repeated during introductory visits by the core team at the 2 participating pain treatment centers. During these visits, the core team also observed multiple treatment sessions and were given a detailed explanation about dose and content of the included treatment modalities. Subsequently, 20 semistructured interviews of approximately 1 hour were conducted and transcribed (see page 4 in Multimedia Appendix 1). During the final activity of phase 1, 60 students divided over 16 teams formulated hypotheses based on the previously collected data and designed provotypes to test their ideas with both healthy participants and patients with chronic pain (see page 5 of Multimedia Appendix 1). At the start of each week, they updated their provotypes based on the received feedback. During the final project session, all teams presented their final provotypes as well as their collected insights to members of the consortium.

Reflection

In phase 1, we were able to create a large qualitative dataset. This dataset not only contained experiences and ideas of stakeholders but also included specific feedback in response to multiple provotypes on a wide array of topics. The consecutive planning of the 3 key activities enabled us to iteratively expand our insights on relapse after pain treatment: Interviews were prepared by using the insights from the kick-off sessions, and the student teams could build upon the preliminary analysis of the available interview data. The participating stakeholders responded positively to the co-design approach and cooperated actively during the sessions and interviews. Despite their inexperience with co-design, the sessions were considered accessible, pleasant, and relevant. However, medical ethical screening procedures and personnel deployment planning limited the possibility for last-minute requests or invitations for including HCPs and patients. The obtained dataset of patient and HCP responses also contributed to a deeper understanding of relevant factors related to relapse, which provided a solid base for further intervention development. For example, the interviews revealed important contextual information such as a “feelings of emptiness after treatment,” “difficulties sharing treatment experiences with friends and family,” and “the different context between the rehabilitation center and the personal environment.”

Phase 2: “Define”

Description

The “Define” phase lasted for 1 month and started with thematically organizing the interviews by means of open coding by the core team (see page 6 of Multimedia Appendix 1) [44,45]. This resulted in 8 main themes and 45 subthemes of factors associated with relapse after successful treatment (see page 7 in Multimedia Appendix 1). To facilitate subsequent co-design activities, the themes were rephrased as questions, plotted on posters, and illustrated with exemplary quotes and figures (see page 8 in Multimedia Appendix 1 for an example). In addition, the core team developed a set of 74 stimulus cards that were designed to facilitate the discussion of specific insights or principles [46]: 36 cards contained insights from the student project, 15 cards contained relevant theory on behavior regulation, and 23 cards contained theory related to chronic pain treatment (see page 9 in Multimedia Appendix 1). Subsequently, patients (4), HCPs (4), researchers (9), designers (6), and students (3) were invited for a co-creation session (see page 10 in Multimedia Appendix 1). During the first assignment, participants were asked to examine the posters and extend them with their own ideas or with stimulus cards. This resulted in 121 notes and 42 cards that were added to the posters. In the second assignment, subgroups were made of participants with varying backgrounds. Each group was instructed to select 1 theme and use the available information to develop an intervention concept. A professional draftsman supported the session by immediately visualizing intervention ideas. The final part of the session consisted of a plenary session where all 5 concepts were presented. During the subsequent discussion, the concepts were compared, and various overarching topics emerged, including “maintaining the positive development after treatment” and “reflection and self-monitoring.” In a subsequent meeting, the steering committee merged these overarching topics into 2 concept ideas: positive reinforcement and direct feedback. The “Define” phase concluded with a design briefing, where the core team commissioned 3 students to develop these ideas into tangible rudimentary prototypes as part of their graduation project.

Reflection

The final system map that included both posters and the card set provided a complete overview of the collected data. This presentation form stimulated participants to combine various insights to develop concept interventions. With respect to stakeholder involvement, the number of patients and HCPs was lower than originally planned. The duration of the session and traveling distance required participants to block a full day, which turned out to be difficult to organize. In line with our findings in phase 1, the co-design methods successfully engaged nonexperts in the design process. The assignment to create concept intervention ideas was concrete and tangible. The resulting 5 concepts were associated with earlier identified patient needs, were grounded in contextual information, and contained relevant insights on relapse prevention. For example, one concept idea focused on monitoring and recognizing early signals of relapse, which was based on stimulus cards (eg, a research insight related to difficulties in unbiased self-monitoring of behavior), interview data (eg, a quote from HCP on the possibility of daily feedback via eHealth), and newly added notes (eg, patient feedback should always be related to patient-specific goals). However, only a fraction of the possible combinations of cards and system maps was explored during this session. Limited time and resources prevented organizing additional sessions to cross-validate the results and achieve saturation.

Phase 3: “Develop”

Description

During the 4 months of the “Develop” phase, students held 5 focus groups to regularly test their ideas with patients and HCPs (see page 11 in Multimedia Appendix 1). For example, by discussing the role of personal values within the treatment program, the students found supporting evidence that these values were strongly related to treatment goals, which subsequently guided the operationalization of the valued-based action plan in one of the rudimentary prototypes. Based on stakeholder feedback and weekly evaluation sessions with the core team, the students worked towards a final rudimentary prototype. One student focused on the transfer of important treatment insights to each patients’ personal context. She developed a toolbox that contained various methods to capture and store therapy insights in order to facilitate retrieval in a relevant personal context (eg, using a personal picture as memory cue for an important moment during treatment). The other 2 students focused on the generation of valued-based goals and the formulation of action plans for each consecutive step towards the goal. Their final rudimentary prototype consisted of a mockup mobile app, allowing participants to browse through all steps that were required to formulate and plan a valued-based goal. In Multimedia Appendix 1, pages 12-13 show the final versions of these rudimentary prototypes.

Reflection

This phase was characterized by a shift from “what” to “how” to design. Accordingly, presentation form, usability, and implementation into existing treatment practice became increasingly relevant. To engage stakeholders, the students visited the treatment centers on multiple occasions. In contrast to other phases, the patients and HCPs could provide feedback on ideas, but were not involved in the decision-making process regarding the final design of the rudimentary prototype, which potentially influenced their commitment. Moreover, their reduced involvement in this phase resulted in limited information regarding the applicability of the rudimentary prototypes in clinical practice.

Phase 4: “Deliver”

Description

In the final phase, the core team merged both rudimentary prototypes into one final prototype intervention over a period of 2 months. To do so, the core team organized a final co-creation session, where the students presented their concepts. The aim of this session was to receive final feedback on the potential value and function of both rudimentary prototypes and to formulate a recommendation to the steering committee with respect to the final prototype design. To facilitate this process, stakeholders (n=14) were instructed to reflect on the concepts by taking various patient perspectives into account. For this purpose, 4 personas were created with variation on 2 characteristics that were often discussed during previous patient interviews. Each persona had either a high or low level of social support and a high or low tendency to protect personal boundaries. In Multimedia Appendix 1, page 14 depicts the discussions between stakeholders as well as the poster that explained the 4 personas. The final conclusion was that both rudimentary prototypes had potential as supportive treatment modalities to prevent relapse after successful treatment. Furthermore, future testing and development should primarily focus on optimizing the active treatment components and calibrating the intervention to the treatment program.

Based on this advice, the steering committee decided to merge both rudimentary prototypes into 1 prototype workbook. The core team composed a list of individual intervention components from each rudimentary prototype (eg, a prompt to set calendar reminders after a goal-setting procedure) and coded these according to the Behaviour Change Technique Taxonomy V1 (see page 15 in Multimedia Appendix 1) [47]. Subsequently, they determined how to transfer the components to a workbook version and performed literature searches to find ideas for optimizing the effectiveness of each component. For example, to assist the formulation of personal values, various value generation procedures were found (eg, [48]) and integrated into the prototype. In addition, the core team checked to which extent the list of intervention components corresponded with the themes of the interview dataset. Of the 19 intervention components, 17 components were related to one or more themes from the dataset, and 27 of the 45 themes were related to one or more intervention components. For example, 4 components in the goal-setting intervention, including specific probing questions to help formulate meaningful values, were associated with the theme “remembering important goals and values after treatment.” A designer, a text editor, and 3 HCPs provided feedback with the conversion to a paper workbook intervention and respectively focused on the design, readability, and appropriate terminology. In Multimedia Appendix 1, page 16 shows examples of the 2 included strategies: the value-based goal forms (b and c) and the Insight Cards (d).

Reflection

Previous difficulties with recruiting sufficient patients for co-creation sessions caused us to search for alternative ways to include their viewpoint. The personas proved a useful method to incorporate various patient perspectives by proxy during the evaluation of the rudimentary prototypes. Furthermore, the validation check indicated that the majority of the intervention components could be traced back to the original stakeholder themes from the interventions in the “Discover” phase and vice versa. This illustrates that stakeholder input has been incorporated in the design. However, the decision to combine both prototype ideas into one intervention was unexpected, which resulted in last-minute planning and consequently in limited stakeholder involvement during the design of the workbook. This may threaten the usability of this prototype in clinical practice.

Discussion

Principal Findings

The primary aim of this study was to reflect on the value and function of co-design methodology during the development of an intervention that prevents relapse after successful pain treatment. In the analysis, we focused on idea generation, stakeholder involvement, and the incorporation of stakeholder input within the development process. Overall, the generative techniques that were employed supported patients and HCPs with sharing their perspectives on pain treatment and relapse, which was in line with our hypothesis. Moreover, the techniques steered the conversations beyond stakeholders' primary responses, often resulting in a detailed account of their personal experiences with the treatment program and of their attempts to integrate treatment insights into their personal environment. In addition, system maps, personas, and prototypes enabled nonexperts to actively participate in design activities. A possible explanation for the successful engagement of stakeholders during the project is that experienced co-designers constantly translated hypotheses and abstract ideas into provotypes or prototypes. This method is particularly useful to provoke user reactions or to rapidly visualize an idea, which evokes interactions with an actual object rather than reflections on past experiences of hypothetical situations [39]. In addition, the used co-design materials helped to transform each location where co-design activities took place (eg, treatment facility or patient home) into a workshop environment that stimulated active participating and emphasized equality between all participants. This is especially important for health care settings, where conventional power relationships between patients and HCPs threaten effective cooperation during design activities [18,19].

With respect to stakeholder involvement, many different patients, HCPs, researchers, students, and designers participated during the study, which was also in line with our hypothesis. The stakeholder interactions mostly consisted of independent design activities that required low commitment and little effort. In contrast, the members of the core team remained active throughout the project, which increasingly created an imbalance in knowledge and involvement between the core team and other participants in co-design activities. This may explain why the role of the stakeholders gradually shifted from “user as partner”—where all participants within the sessions contributed as equals in the design activities—towards “users as subject”—where participants mainly provided expert opinions or performed delimited tasks (eg, usability testing) [49]. Consequently, the concepts underlying the intervention have been thoroughly grounded in stakeholder input and expertise, but the applicability of the current workbook operationalization within the treatment programs requires further testing to examine whether the current strategies fit patient preferences and can be integrated in treatment programs in the form of the current prototype.

This project shows similarities to the experience-based co-design (EBCD) approach, which aims to improve health care services by actively involving stakeholders to collect knowledge and experiences, to set priorities, and to develop solutions [18,50]. Although this project did not follow the 6 stages of EBCD, the overall objectives as well as the systematic partnership with patients, HCPs, designers, and researchers are alike. A notable difference was the focus within this project on actual prototype development throughout all phases, which promoted a solution-focused orientation for the participants. Alternatively, in EBCD, more emphasis is placed on ensuring that the collected patient experiences are received and understood by other stakeholders (eg, by showing a film of patient interview segments that reflect key themes), before continuing to developing improvements [18]. These differences illustrate the versatility of co-design and its potential to adapt to different design environments.

Strengths and Limitations

The extensive documentation of the co-design activities allowed for a detailed reconstruction of the development process. Furthermore, during co-creation sessions, steering committee meetings, and the construction of the retrospective journey, representatives from all research groups were present, which resulted in a continuous integration of various perspectives during the project. However, we did not film or record any of the co-creation sessions. Although analyzing audio or video would have been time consuming, it would have provided further possibilities to observe stakeholder discussions during design activities and to include additional insights that we did not record.

During the project, we experienced a tradeoff between validating the outcomes of co-design activities and analyzing the results for the next iteration. For example, an additional co-creation session during the "Define" phase with different stakeholders could have cross-validated the outcomes of the initial session. However, given limited resources, this would have resulted in fewer development iterations in the remaining period. A key argument in favor of more iterations is that quickly integrating stakeholder input into subsequent sessions directly visualizes the value of their input [51]. However, a tendency towards more iterations increases the uncertainty to what extent the outcomes of this project can be generalized to the population [33].

Future Recommendations

This study adds to the increasing number of initiatives that use co-design to structurally integrate contextual factors into the development of health care interventions (eg, [21,22,52,53]), which help bridge the gap from development to actual implementation [12,13]. When using co-design, it is important to relate the findings of the process to existing theories and treatments, for instance by using the behavior change technique taxonomy [23,47]. This strengthens the co-design approach by combining stakeholder evaluations with existing theory. Importantly, further integration between co-design and theory-driven approaches only becomes possible when using rigorous testing to evaluate the outcomes of the co-design process [13]. Consequently, an important next step in answering the question about whether co-design helps inform the development of health interventions will involve more examples of development projects. In these examples, co-design–based interventions are ideally subjected to experimental testing. Furthermore, we believe that future co-design projects in the health care domain should include detailed planning of activities and take lengthy medical ethical approval procedures into account [51].

Conclusions

To acquire a better understanding of how co-design may benefit the development of interventions in the health care domain, examples of good practice are necessary**.** In this article, we presented such an example. By critically reviewing the value and function of a co-design project with respect to idea generation, stakeholder involvement, and incorporation of stakeholder input into the development process, we demonstrated how co-design contributed to the transition from ideas, via concepts, to prototypes.

Acknowledgments

We thank the researchers Jan Pool, Manon Beetstra-Huszar, Albère Köke, and Roos Tigchelaar for their continuous support with collecting and analyzing the data. We also thank Rianne Schaekens, Deloryan Hommers, and Yasmijn Baas for their assistance in the development phase during their graduation projects.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the co-design development process.
[PDF File (Adobe PDF File), 16783 KB - jmir_v23i1e18462_app1.pdf]

References

1. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. BMJ 1998 Aug 15;317(7156):465-468 [FREE Full text] [doi: 10.1136/bmj.317.7156.465] [Medline: 9703533]
2. Chambers CT. From evidence to influence: dissemination and implementation of scientific knowledge for improved pain research and management. Pain 2018 Sep;159 Suppl 1:S56-S64. [doi: 10.1097/j.pain.0000000000001327] [Medline: 30113948]
3. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
4. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
5. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
6. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
7. Kok G, Schaalma H, Ruiter RAC, van Empelen P, Brug J. Intervention mapping: protocol for applying health psychology theory to prevention programmes. J Health Psychol 2004 Jan;9(1):85-98. [doi: 10.1177/1359105304038379] [Medline: 14683571]
8. Sheeran P, Klein WMP, Rothman AJ. Health Behavior Change: Moving from Observation to Intervention. Annu Rev Psychol 2017 Jan 03;68:573-600. [doi: 10.1146/annurev-psych-010416-044007] [Medline: 27618942]
9. Bishop FL, Fenge-Davies AL, Kirby S, Geraghty AWA. Context effects and behaviour change techniques in randomised trials: a systematic review using the example of trials to increase adherence to physical activity in musculoskeletal pain. Psychol Health 2015 Jan;30(1):104-121. [doi: 10.1080/08870446.2014.953529] [Medline: 25109300]
10. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
11. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for Sustained Use of an Activity Tracker: Observational Study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
12. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
13. Robert G, Macdonald AS. Co-design, organisational creativity and quality improvement in the healthcare sector: ‘designerly’ or ‘design-like’? In: Sangiorgi D, Prendville A, editors. Design for service: Key issues and new directions. London, England: Bloomsbury Academic; 2017:117-130.
14. Tarquinio C, Kivits J, Minary L, Coste J, Alla F. Evaluating complex interventions: perspectives and issues for health behaviour change interventions. Psychol Health 2015 Jan;30(1):35-51. [doi: 10.1080/08870446.2014.953530] [Medline: 25140439]
15. Stappers PJ, Giaccardi E. Research through design. The encyclopedia of human-computer interaction.: The Interaction Design Foundation; 2017. URL: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design [accessed 2020-12-15]
16. Steen MGD. Thesis: The fragility of human-centred design. Delft University of Technology. 2008. URL: http://resolver.tudelft.nl/uuid:2b930c09-0ad3-41f8-93f0-503eaea04ff0 [accessed 2020-12-15]
17. Bradwell P, Marr S. Making the most of collaboration an international survey of public service co-design. Annual Review of Policy Design. 2017. URL: https://www.demos.co.uk/files/CollabWeb.pdf [accessed 2020-12-15]
18. Donetto S, Pierri P, Tsianakas V, Robert G. Experience-based Co-design and Healthcare Improvement: Realizing Participatory Design in the Public Sector. The Design Journal 2015 May 07;18(2):227-248. [doi: 10.2752/175630615x14212498964312]
19. Boyd H, McKernon S, Mullin B, Old A. Improving healthcare through the use of co-design. N Z Med J 2012 Jun 29;125(1357):76-87. [Medline: 22854362]
20. Iedema R, Merrick E, Piper D, Britton K, Gray J, Verma R, et al. Codesigning as a Discursive Practice in Emergency Health Services: The Architecture of Deliberation. The Journal of Applied Behavioral Science 2010 Mar 31;46(1):73-91. [doi: 10.1177/0021886309357544]
21. Jamin G, Luyten T, Delsing R, Braun S. The process of co-creating the interface for VENSTER, an interactive artwork for nursing home residents with dementia. Disabil Rehabil Assist Technol 2018 Nov;13(8):809-818. [doi: 10.1080/17483107.2017.1385102] [Medline: 29037109]
22. Revenäs Å, Martin C, H Opava C, Brusewitz M, Keller C, Åsenlöf P. A Mobile Internet Service for Self-Management of Physical Activity in People With Rheumatoid Arthritis: Challenges in Advancing the Co-Design Process During the

Requirements Specification Phase. JMIR Res Protoc 2015 Sep 17;4(3):e111 [FREE Full text] [doi: 10.2196/resprot.4824] [Medline: 26381221]

23. Verbiest MEA, Corrigan C, Dalhousie S, Firestone R, Funaki T, Goodwin D, et al. Using codesign to develop a culturally tailored, behavior change mHealth intervention for indigenous and other priority communities: A case study in New Zealand. Transl Behav Med 2019 Jul 16;9(4):720-736. [doi: 10.1093/tbm/iby093] [Medline: 30388262]
24. Greger S, Hatami Z. Reducing social distance through co-design. In: Keinonen T, Vaajakallio K, Honkonen J, editors. Designing for wellbeing. Helsinki, Finland: Aalto Arts Books; 2013.
25. Morley S. Relapse prevention: still neglected after all these years. Pain 2008 Feb;134(3):239-240. [doi: 10.1016/j.pain.2007.12.004] [Medline: 18164134]
26. Turk DC, Rudy TE. Neglected topics in the treatment of chronic pain patients--relapse, noncompliance, and adherence enhancement. Pain 1991 Jan;44(1):5-28. [doi: 10.1016/0304-3959(91)90142-k] [Medline: 2038489]
27. Barab SA. Design-based research: A methodological toolkit for engineering change. In: Sawyer K, editor. Handbook of the Learning Sciences, Vol 2. Cambridge, MA: Cambridge University Press; 2014:233-270.
28. Toye F, Seers K, Hannink E, Barker K. A mega-ethnography of eleven qualitative evidence syntheses exploring the experience of living with chronic non-malignant pain. BMC Med Res Methodol 2017 Aug 01;17(1):116 [FREE Full text] [doi: 10.1186/s12874-017-0392-7] [Medline: 28764666]
29. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth Delivered Interventions: A Systematic Review to Assess Key Methods and Processes. Curr Nutr Rep 2016 Jul 4;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
30. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650 [FREE Full text] [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
31. Yin RK. Case study research design and methods, 3rd edition. Thousand Oaks, CA: Sage Publications, Inc; 2003.
32. Foth M, Axup J. Participatory designaction research: Identical twins or synergetic pair? In: Wagner I, Jacucci G, Kensing F, Blomberg J, editors. Expanding Boundaries in Design: Proceedings Ninth Participatory Design Conference 2006 (Vol 2). Canada: Computer Professionals for Social Responsibility; 2006:93-96.
33. Seekins T, White GW. Participatory action research designs in applied disability and rehabilitation science: protecting against threats to social validity. Arch Phys Med Rehabil 2013 Jan;94(1 Suppl):S20-S29. [doi: 10.1016/j.apmr.2012.07.033] [Medline: 23178031]
34. White GW, Suchowierska M, Campbell M. Developing and systematically implementing participatory action research. Arch Phys Med Rehabil 2004 Apr;85(4 Suppl 2):S3-12. [doi: 10.1016/j.apmr.2003.08.109] [Medline: 15083417]
35. Howard Z, Somerville MM. A comparative study of two design charrettes: implications for codesign and participatory action research. CoDesign 2014 Mar 06;10(1):46-62. [doi: 10.1080/15710882.2014.881883]
36. Visser FS, Stappers PJ, van der Lugt R, Sanders EB. Contextmapping: experiences from practice. CoDesign 2005 Apr;1(2):119-149. [doi: 10.1080/15710880500135987]
37. Howard T. Journey mapping. Commun. Des. Q. Rev 2014 May;2(3):10-13. [doi: 10.1145/2644448.2644451]
38. Trebble TM, Hansi N, Hydes T, Smith MA, Baker M. Process mapping the patient journey: an introduction. BMJ 2010 Aug 13;341:c4078. [doi: 10.1136/bmj.c4078] [Medline: 20709715]
39. Sanders EB, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014 Mar 06;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
40. Boer L, Donovan J. Provotypes for participatory innovation. In: DIS '12: Proceedings of the Designing Interactive Systems Conference. New York, NY: Association for Computing Machinery; 2012:388-397.
41. Jones P. Design for care: Innovating healthcare experience. New York, NY: Rosenfeld Media; 2013.
42. Adlin T, Pruitt J, Goodwin K, Hynes C, McGrane K, Rosenstein A, et al. Putting personas to work. In: CHI'06 Extended Abstracts on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2006:13-16.
43. The Design Process: What is the Double Diamond? Design Council. 2015. URL: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond [accessed 2020-12-15] [WebCite Cache ID 74d5rk9mE]
44. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
45. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
46. Coughlan P, Suri JF, Canales K. Prototypes as (Design) Tools for Behavioral and Organizational Change. The Journal of Applied Behavioral Science 2016 Jul 26;43(1):122-134. [doi: 10.1177/0021886306297722]
47. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

48. Chase JA, Houmanfar R, Hayes SC, Ward TA, Vilardaga JP, Follette V. Values are not just goals: Online ACT-based values training adds to goal setting in improving undergraduate college student performance. Journal of Contextual Behavioral Science 2013 Oct;2(3-4):79-84. [doi: 10.1016/j.jcbs.2013.08.002]
49. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
50. Robert G, Cornwell J, Locock L, Purushotham A, Sturmey G, Gager M. Patients and staff as codesigners of healthcare services. BMJ 2015 Feb 10;350:g7714. [doi: 10.1136/bmj.g7714] [Medline: 25670179]
51. Hinchcliff R, Greenfield D, Braithwaite J. Is it worth engaging in multi-stakeholder health services research collaborations? Reflections on key benefits, challenges and enabling mechanisms. Int J Qual Health Care 2014 Apr;26(2):124-128. [doi: 10.1093/intqhc/mzu009] [Medline: 24519121]
52. O'Brien N, Heaven B, Teal G, Evans EH, Cleland C, Moffatt S, et al. Integrating Evidence From Systematic Reviews, Qualitative Research, and Expert Knowledge Using Co-Design Techniques to Develop a Web-Based Intervention for People in the Retirement Transition. J Med Internet Res 2016 Aug 03;18(8):e210 [FREE Full text] [doi: 10.2196/jmir.5790] [Medline: 27489143]
53. Raynor DK, Ismail H, Blenkinsopp A, Fylan B, Armitage G, Silcock J. Experience-based co-design-Adapting the method for a researcher-initiated study in a multi-site setting. Health Expect 2020 Jun 11;23(3):562-570 [FREE Full text] [doi: 10.1111/hex.13028] [Medline: 32045087]

Abbreviations

EBCD: experience-based co-design
HCP: health care provider
PAR: participatory action research

Edited by G Eysenbach; submitted 27.02.20; peer-reviewed by M Bestek, S Berrouiguet, I Shubina; comments to author 20.04.20; revised version received 16.06.20; accepted 23.11.20; published 20.01.21.

Please cite as:
Elbers S, van Gessel C, Renes RJ, van der Lugt R, Wittink H, Hermsen S
Innovation in Pain Rehabilitation Using Co-Design Methods During the Development of a Relapse Prevention Intervention: Case Study
J Med Internet Res 2021;23(1):e18462
URL: http://www.jmir.org/2021/1/e18462/
doi:10.2196/18462
PMID:33470937

Original Paper

Ride-Hailing Services and Alcohol Consumption: Longitudinal Analysis

Gordon Burtch[1*], PhD; Brad N Greenwood[2*], PhD; Jeffrey S McCullough[3*], PhD

[1]Carlson School, University of Minnesota, Minneapolis, MN, United States
[2]Department of Information Systems and Operations Management, George Mason University, Fairfax, VA, United States
[3]Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, United States
[*]all authors contributed equally

Corresponding Author:
Jeffrey S McCullough, PhD
Department of Health Management and Policy
University of Michigan
1415 Washington Heights
Ann Arbor, MI
United States
Phone: 1 734 936 1189
Fax: 1 734 764 4338
Email: jemccull@umich.edu

Abstract

Background: Alcohol consumption is associated with a wide range of adverse health consequences and a leading cause of preventable deaths. Ride-hailing services such as Uber have been found to prevent alcohol-related motor vehicle fatalities. These services may, however, facilitate alcohol consumption generally and binge drinking in particular.

Objective: The goal of the research is to measure the impact of ride-hailing services on the extent and intensity of alcohol consumption. We allow these associations to depend on population density as the use of ride-hailing services varies across markets.

Methods: We exploit the phased rollout of the ride-hailing platform Uber using a difference-in-differences approach. We use this variation to measure changes in alcohol consumption among a local population following Uber's entry. Data are drawn from Uber press releases to capture platform entry and the Behavioral Risk Factor Surveillance Systems (BRFSS) Annual Survey to measure alcohol consumption in 113 metropolitan areas. Models are estimated using fixed-effects Poisson regression. Pre- and postentry trends are used to validate this approach.

Results: Ride-hailing has no association with the extent of alcohol consumption in high (0.61 [95% CI –0.05% to 1.28%]) or low (0.61 [95% CI –0.05% to 1.28%]) density markets, but is associated with increases in the binge drinking rate in high-density markets (0.71 [95% CI 0.13% to 1.29%]). This corresponds to a 4% increase in binge drinking within a Metropolitan Statistical Area.

Conclusions: Ride-hailing services are associated with an increase in binge drinking, which has been associated with a wide array of adverse health outcomes. Drunk driving rates have fallen for more than a decade, while binge drinking continues to climb. Both trends may be accelerated by ride-hailing services. This suggests that health information messaging should increase emphasis on the direct dangers of alcohol consumption and binge drinking.

(J Med Internet Res 2021;23(1):e15402) doi:10.2196/15402

KEYWORDS

binge drinking; drunk driving; road traffic safety; ride-hailing; alcohol consumption; difference in differences; Uber

Introduction

Background

Alcohol consumption is a serious public health issue with significant implications for personal health and well-being [1]. Between 2006 and 2010, more than 88,000 people lost their lives as a result of alcohol abuse in the United States; 56% of those deaths occurred because of acute events like motor vehicle collisions or alcohol poisoning [2]. Scholars have tied alcohol consumption to a host of deleterious societal outcomes, including spousal and child abuse [3]; inability to maintain gainful employment [4]; and personal health concerns in the form of diabetes, liver disease, and sexual dysfunction [5]. Prior work has also documented many associated factors that can accelerate alcohol consumption, including peer pressure and social norms [6], financial distress [7], and mobility and public transit accessibility [8], the context of this investigation.

Urban mobility is particularly notable when one considers the changes in the market for individual transportation over the last 10 years (ie, ride-hailing). Whereas people were once forced to drive themselves or depend on often unreliable alternatives (eg, public transit or taxis), ride-hailing services now offer a simple solution that is integrated with an individual’s smartphone. Uber, for example, completed its 10 billionth ride in June 2018, and ride-hailing more generally accounts for a significant share of the urban mobility market. Moreover, recent academic work has begun to assess the relationships between ride-hailing services and public health outcomes (eg, traffic fatalities [9,10], occurrence of assault [11-13], ambulance use [14], and rates of drunk driving [15]). In this work, we delve into the question of whether ride-hailing services affect the extent of alcohol consumption within the population (ie, prevalence) as well as the intensity of consumption (binge drinking).

Research Objectives

Research into the social implications of ride-hailing services is wide ranging. A significant body of work has been devoted to the economic implications of ride-hailing, including the effect on entrepreneurship [16], durable goods purchases [17], and the labor market more broadly [18]. Research has also begun to examine social issues stemming from ride-sharing. Findings are diverse and include diminished rates of sexual assault [12], increases in property crimes [11], and a reemergence of bias based on ascriptive characteristics which has traditionally be absent from online transactions [19,20]. However, the largest body of work in this space has, unsurprisingly, focused on motor vehicle safety [9-11] and, more specifically, the effect on drunk driving [13,15].

The potential drunk driving and ride-hailing relationship is intuitive. Platforms like Uber are more accessible than limousine services and more ubiquitous and reliable than taxis [21,22], and thus people may shift from driving under the influence to ride-hailing. The fact that ride-hailing is often a complement to public transportation [23,24] only underscores this idea. Because ride-hailing services can solve the last mile problem for people reliant on public transportation and provide point-to-point transportation for those not reliant on public transportation, the need to drive should be diminished.

We build on past lines of inquiry and offer a first consideration of whether ride-hailing services facilitate a greater extent of alcohol consumption (a larger number of individuals consuming alcohol) and a greater intensity of alcohol consumption (more binge drinking). Ride-hailing services might simply displace alcohol consumption which used to occur at home by making it easier to go to bars, restaurants, and night clubs. However, past work has found that people elevate their alcohol consumption in the presence of an assured transportation option (eg, a designated driver [25]). For example, when the city of Washington, DC, extended the Washington Metropolitan Area Transit Authority train hours (ie, mass transit) to after last call, there was not only a sharp drop in drunk driving but a significant increase in property crimes associated with drinking to excess [8]. Given the relative ease with which individuals can obtain a ride-hailing trip following a night of drinking, it is plausible that ride-hailing services would increase both the extent and intensity of alcohol consumption. Coupled with the public costs associated with excessive alcohol consumption, approximately $250 billion in the United States in 2010 [26], any connection between ride-hailing access and drinking activity would have important policy implications.

Our analyses test two hypotheses. Our first hypothesis (H1) is that access to ride-hailing services increases alcohol consumption at the extensive margin (ie, raising the number of individuals drinking alcohol). Our second hypothesis (H2) is that access to ride-hailing services increases alcohol consumption at the intensive margin (ie, the amount of alcohol a given individual consumes in a sitting). The distinction between these hypotheses is important. While public health interventions and many laws focus on the intensity of individual alcohol consumption (drinking to excess), recent work suggests that the extent of alcohol consumption may also be a concern in its own right as even one drink of alcohol daily has been associated with negative health outcomes [1].

Considering our two hypotheses, it is also important to note previous research, which suggests that the effects of ride-hailing services are heterogeneous across markets [23,24]. Ride-hailing platforms are subject to network effects (riders require a sufficient supply of drivers and vice versa), thus population density is typically crucial to their adoption. Further, past work notes that ride-hailing services are often employed as complements to public transit [23], and thus their relationship to drinking activity is likely to be more apparent in locations characterized by higher population density, where transit options are more readily available (eg, taking the train into the city and an Uber home). Finally, locations characterized by greater population density are also more likely to bear a greater density of drinking establishments, suggesting a larger upper bound on the possible rise in drinking. Given these points, we conduct our analyses while distinguishing between locations characterized by higher versus lower population density.

Methods

Data

We combine data on the diffusion of ride-hailing services with information on alcohol consumption and population density.

Data on alcohol consumption are drawn from the Behavioral Risk Factors Surveillance Survey (BRFSS). The BRFSS data are collected as part of the Selected Metropolitan Area Risk Trends program from the Centers for Disease Control and Prevention (CDC). These data include annual counts of respondent answers to questions about risky behaviors, including alcohol consumption, for a select set of metropolitan areas. We examine two such questions: respondents' engaging in any drinking in the prior 30 days and binge drinking in the past week (the latter defined as consuming 5 or more drinks in a single session for males and 4 or more drinks for females). BRFSS measures for any and binge drinking are not available for every Metropolitan Statistical Area (MSA) year because the survey is conducted by the CDC only in those locations that meet certain sampling criteria, and positive response counts (responses of yes) are reported only in cases where a minimum threshold of 50 responses was obtained for the MSA. As a result, some locations enter and exit the sample over time. That said, numerous studies have been conducted that speak to the validity and representativeness of the BRFSS data.

Data on Uber's entry into different locations are drawn from public press releases and announcements on Uber's website (see Multimedia Appendix 1 for a list of Uber entry dates for the geographies comprising our sample). Uber presence is coded as a dichotomous indicator equal to 1 once Uber (in any format, whether Uber X or Uber Black) enters any city or township within a given MSA, and 0 in preceding years. We focus our analysis on Uber for 3 reasons. First, competing services (eg, Lyft) systematically enter after Uber, except in a handful of rare instances. Second, Uber is the largest ride-hailing platform by rider base and revenue ($82.4 billion in 2019) [27]. Third, the other dominant ride-hailing service in the United States, Lyft, operates a very similar business model, implying it would have a very similar effect. In fact, they are so similar that drivers and riders frequently multihome, switching between the two services [28].

Finally, we incorporate US Census Bureau measures of MSA population density, because prior work has documented that Uber's influence is particularly pronounced in dense urban areas [15]. Markets with dense populations are thicker—having more potential drivers, consumers, and destinations within a region—and it is reasonable to expect that consumers experience better service with reduced wait times and lower fares. As such, we expect larger associations in these population dense areas.

Our sample spans 2010 through 2016, comprising 113 MSAs in which both ride-hailing entry dates were recoverable, and where BRFSS measures of alcohol consumption were available for at least 2 years. Each MSA is observed for an average of 6.5 years. This combined sample, coupled with the fact that Uber enters different locations at different times, allows us to measure systematic changes in alcohol consumption after ride-hailing service entry. Summary statistics are in Table 1. As can be seen in the table, any drinking is reported by 56.12% (1025.78/1827.91) of respondents on average, while 17.23% (315.03/1827.91) report binge drinking. Uber first enters our sample in 2011. Entry was gradual for the first few years, accelerated rapidly in 2014, and was nearly universal by 2016.

Table 1. Descriptive statistics.

Variable	Mean (SD)	Median	Minimum	Maximum	Observations
Uber presence	0.41 (0.49)	0	0	1	730
Binge drinking[a]	315.03 (311.67)	205	18	2067	728
Any drinking[b]	1025.78 (1016.11)	675	79	6030	730
BRFSS responses[c]	1827.91 (1682.07)	1293.5	464	9333	730
Population density[d]	2469.02 (2191.48)	1953.5	7.2	13,597.50	588

[a] Binge drinking reflect the number of respondents who indicated they had 5 or more drinks (in the case of males; 4 or more drinks in the case of females,)on a single occasion in the prior 30 days.

[b] Any drinking reflects the number of respondents indicating they consumed at least one alcoholic beverage in the prior 30 days.

[c] BRFSS: Behavioral Risk Factors Surveillance Survey.

[d] Population density reflects people per square mile within 10 miles of city hall.

Statistical Analysis

Measuring the relationship between alcohol consumption and ride-hailing service availability is inherently challenging, particularly given that Uber may have selectively entered markets characterized by heavy—or even growing—rates of alcohol consumption in an effort to optimize both its ridership and profits. That is, ride-hailing services are known to be used more intensely on Friday and Saturday nights, and Uber may have selectively entered markets with higher levels of alcohol consumption with the objective of capturing that demand. We address this selection problem by employing a difference-in-differences approach [29,30]. In doing so, we measure the change in alcohol consumption before and after Uber entered into an MSA relative to the change in markets where Uber had yet to enter. Parameters are estimated based on a difference-in-differences Poisson regression, incorporating location (α_l) and time (τ_t) fixed effects, with standard errors clustered by MSA. Equation 1 reflects the regression model that we estimate. Our independent variable of interest is *Uber*, which takes a value of 1 if MSA l had received any form of Uber as of year t. Our coefficient of interest is β, which reflects our estimate of the relationship between Uber entry and alcohol consumption. Additionally, we incorporate BRFSS respondent counts for MSA l as an exposure term in the regression.

$$Log\ (Drinking_{l,t}) \sim \alpha_l + \tau_t + \beta \cdot Uber_{l,t} + \varepsilon_{l,t}\ (1)$$

We calculate and report the estimated change in the incidence rate of any drinking and binge drinking among respondents in a location based on marginal effects. Further, given that ride-hailing's impact should differ across markets based on population density, we split the sample and estimate separate models for markets having above- and below-median population densities, based on measures reported by the US Census Bureau, derived from the 2010 Census.

Robustness and Sensitivity

Before examining the results of the above estimations, we first discuss the robustness and sensitivity checks. First, we began by testing for pre-entry changes in alcohol consumption in the periods leading up to Uber's entry into an MSA. Significant differences prior to Uber's arrival would imply violation of the parallel trend assumption that underlies difference-in-differences estimation. Evaluating the parallel trends assumption is important to ensure our estimates are not driven by Uber systematically selecting into cities, in a manner that merely correlates with expected or ongoing growth in drinking activity. As Uber representatives have emphasized its widespread use in traveling to or from bars on weekend nights, selection effects on Uber's part are certainly possible and of concern. Put another way, while the location fixed effects included in our regression model make it robust to Uber entering markets with higher persistent levels of drinking activity or any other stable and unchanging features of that location, systematic entry by Uber into markets that exhibit growing trends in binge drinking, even before Uber's arrival, would violate the baseline assumptions of the model by creating a dynamic, time-varying, confounder for the estimated association between Uber presence and alcohol consumption.

To implement this test, we estimate a relative time variant of difference-in-differences regression [30]. In doing so, we construct a set of indicators capturing the relative (to Uber entry) year that an observation had taken place, which is defined based on the timing of Uber's arrival in a particular MSA. In this regression, we omit the 2 years preceding Uber's arrival, taking these as a joint reference period. It should be noted that a degree of freedom is lost in the estimation because we jointly estimate both absolute and relative time dummies (ie, year and year from Uber entry), so two reference periods must be omitted for the other model parameters to be identified. This specification allows us to estimate the relationship between Uber entry and drinking flexibly, as we can observe differences (or the absence of difference) in both pre- and postentry drinking trends. With this regression, if the assumptions of the model hold, we expect to observe positive and significant differences in the years following Uber's arrival relative to the 2 years just prior to an Uber arrival in any MSA but no significant differences in the years prior (eg, 3, 4, or 5 years prior).

Second, we explored sensitivity of our results to our choice of population density measure. Our baseline analyses employ a measure from the Census Bureau reflecting the population per square mile residing within 10 miles of city hall. Given that this choice of radius is to some degree arbitrary, our goal is to ensure that any results are not sensitive to the choice. We therefore also considered splits using alternative measures of population per square mile, namely with a radius of 5 miles to city hall and again within 2 miles of city hall.

Results

Statistical Analysis

Results of our primary regression analyses appear in Table 2. We observe no significant relationship between Uber's presence and the extent of alcohol consumption (ie, any drinking). Table 2 indicate a lack of significant change in any drinking across both high- and low-density markets. Uber's entry is, however, associated with a significant rise in binge drinking within high-density MSAs. Table 2 reports a marginal estimate of 0.71 (95% CI 0.133% to 1.289%). This indicates that Uber's entry is associated with an additional ~0.70% increase in the population's binge drinking rate. This corresponds to a 4% relative increase in the binge drinking population in a given MSA.

Table 2. Marginal effects of ride-hailing on alcohol consumption in high- and low-density Metropolitan Statistical Areas.

Characteristic	Uber presence % (SE)	95% CI	Year fixed effects	MSA[a] fixed effects	Observations	MSAs	χ^2
Any drinking							
Low density	–1.10 (0.63)	–2.33 to 0.13	Yes	Yes	264	61	64.99 (7)
High density	0.61 (0.34)	–0.05 to 1.28	Yes	Yes	324	52	120.15 (7)
Binge drinking							
Low density	–0.12 (0.61)	–1.31 to 1.07	Yes	Yes	263	61	700.57 (7)
High density	0.71 (0.30)	0.13 to 1.29	Yes	Yes	324	52	265.65 (7)

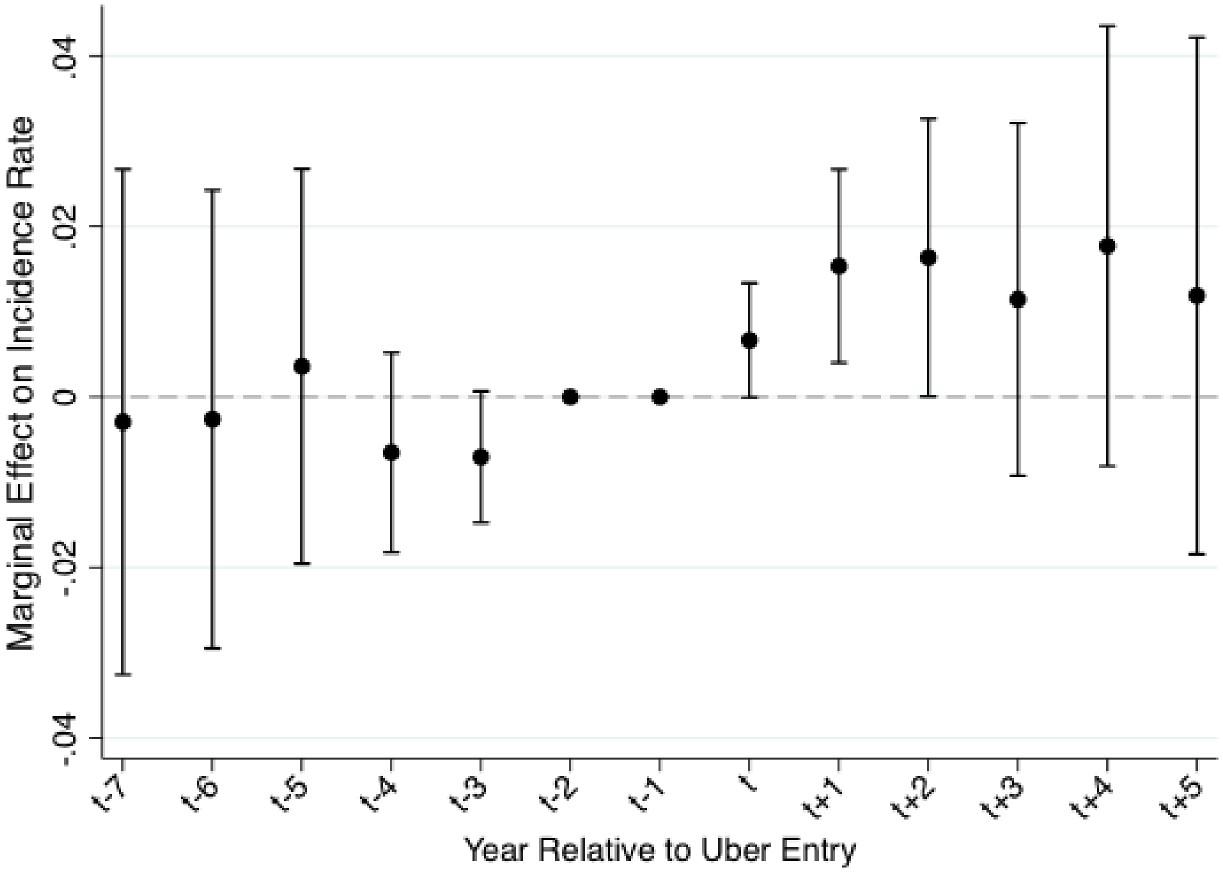
[b]MSA: Metropolitan Statistical Area.

It should be noted that the estimated association between Uber presence and binge drinking in high-density MSAs persists following a correction for multiple comparisons. Specifically, applying a very conservative Bonferroni correction, we would rely on a significance threshold of 0.025 (for a 1-tailed test, consistent with the directional nature of the hypotheses). The observed *P* value associated with the relationship between Uber presence and binge drinking within high-density MSAs is .02. Although we observe a statistically significant association between Uber presence and binge drinking in high-density MSAs, it is important to note that the estimated coefficient is not significantly different from the estimate recovered in low-density MSAs. Thus, our findings only support a conclusion that Uber is significantly associated with a rise in binge drinking in high-density MSAs, not that the association is systematically larger in high-density MSAs as compared with low-density MSAs [31].

Robustness and Sensitivity Results

Figure 1 provides a graphical representation of relative time differences for binge drinking in high-density MSAs. As discussed, we employed this model to test for time-varying differences in binge drinking preceding Uber's entry, which would indicate a possible spurious relationship. Each point is a parameter estimate corresponding to leading (t–7 to t–3) and lagged (t to t+5) differences. Consistent with prior work, all estimates are conducted relative to the 2 years (t–1 and t–2) preceding Uber's entry (period t) [15]. The leading, or preperiod, estimates are uniformly small and statistically insignificant despite our use of relatively conservative (90%) confidence intervals. These results provide empirical support for our difference-in-differences strategy as they suggest no violation of the parallel trends assumption. It therefore appears reasonable for us to believe that Uber's entry is exogenous with respect to drinking activity, conditional upon controls. The significant association with binge drinking persists in years following Uber's entry into a market (ie, t+1 to t+5). These findings are again consistent with our difference in differences strategy. As expected, the confidence intervals are wider in the tails of our distribution (eg, for t–4 or less and t+3 or more) where there is less data available to perform the estimation (ie, there is a lack of power in the estimations because few cities have had Uber for such a lengthy period of time). In sum, our findings are consistent with those reported in Table 2 and provide empirical support for the assumptions of our research design (ie, parallel trends).

Figure 1. Marginal effects of relative year dummies in dense Metropolitan Statistical Areas (90% confidence intervals).



Results of our sensitivity around the population density measure are in Table 3 (ie, population per square mile within 5 miles and 2 miles of city hall). With each alternative measure, we continue to observe that higher density MSAs exhibit a significant increase in the incidence of binge drinking following Uber's arrival. No significant association is observed in low density MSAs.

Table 3. Robustness check: marginal binge drinking effects using alternative density measures.

Characteristic	Uber presence % (SE)	95% CI	Year fixed effects	MSA[a] fixed effects	Observations	MSAs	χ^2
5 miles of city hall							
Low	–0.65 (0.44)	–1.50 to 0.21	Yes	Yes	322	74	125.50 (7)
High	0.63 (0.23)	0.17 to 1.09	Yes	Yes	406	74	345.94 (7)
2 miles of city hall							
Low	–0.14 (0.56)	–1.24 to 0.96	Yes	Yes	317	71	108.53 (7)
High	0.52 (0.24)	0.06 to 0.99	Yes	Yes	406	72	346.20 (7)

[a] MSA: Metropolitan Statistical Area.

Discussion

Principal Findings

We examined the association between ride-hailing platform entry and alcohol consumption. While prior work has tied ride-hailing to decreases in levels of drunk driving [13,15], the secondary implications of low-cost urban mobility have received scant attention. Results are nuanced. On one hand, ride-hailing is not associated with the number of people who consume alcohol in general (ie, the extensive margin). This is encouraging and suggests the presence of ride-hailing is not causing greater numbers of people to consume alcohol. The platform is, however, significantly associated with binge drinking in densely populated markets (ie, the intensive margin). This is cause for concern. Given the national and personal costs associated with binge drinking [32], these findings compound an already alarming national trend [33]. This suggests that although ride-hailing may reduce drunk driving, people update their behavior to exploit the assured mobility ride-hailing offers.

These findings suggest that Uber's entry increases binge drinking by 4% in adult urban populations and point to valuable directions for future research. Prior work indicates that most binge drinking occurs among the young (aged 18 to 34 years) and is twice as common among men [25]. As ride-hailing associates with binge drinking, it is likely that the association would be concentrated in younger populations. However, this is speculative, and it is important to assess where behavioral changes are occurring. To the extent that means of viable travel are already established for younger groups, specifically when they are under the influence, it is plausible that these groups may also fail to exhibit changes in their behavior once Uber arrives. Ride-hailing may, instead, be associated with binge drinking increases among groups that are traditionally less at risk, by affording them greater mobility. It is also important to assess whether or not these differences manifest asymmetrically between men and women. Inasmuch as alcohol consumption has been associated with the incidence of sexual assault [12,34] and lesser infractions like property crime [8], the possible negative implications of Uber arrival for public safety through its relationship with binge drinking are important to consider. These services may also benefit public safety (eg, by affording a secure means of transport home late at night). Additional work is clearly needed to tease out whether ride-hailing services have a net beneficial or detrimental association with public safety measures.

Limitations

This work is subject to certain limitations. Most notably, our estimates focus on the entry of ride-hailing services. The exact reasons that ride-hailing services significantly associate with increases in binge drinking in our sample are not altogether clear. This is a natural limitation of secondary data. Several possibilities exist. For example, recent work has noted that ride-hailing services can, in some instances, enhance access to public transit [23]. As such, the result we observe may be explained not just by ride-hailing services access but access in tandem with public transit options (which tend to be better in areas characterized by high population density).

Because we lack individual level data, the underlying associations are likely to be much larger for subpopulations that are most affected. One potentially valuable path forward, which our sample does not allow us to explore, are to consider effects across different subpopulations (eg, age, gender, and local area characteristics like educational composition and access to mass transit). In this context, where we to explore various sources of heterogeneity in a data-driven, exploratory manner, it would be necessary to implement stringent corrections for multiple corrections effectively reducing our already limited statistical power. Future work can look to address these aspects.

Our findings may be vulnerable to time-varying unobserved confounds (ie, unobserved factors) that might correlate over time with both Uber's entry and binge drinking rates. That said, for this to be a concern, the dynamic confounds would need to vary systematically with Uber's entry timing, which seems unlikely. Finally, our sample is potentially not representative, given the CDC's application of sampling criteria in its administration of the BRFSS. The fact that the BRFSS does not report levels of drinking for all MSA-years does raise concerns of external validity. Yet, as the reported MSAs are almost uniformly larger and more densely populated, it is likely that the unreported MSAs would react in a way that is similar to the less densely populated MSAs reported in the sample. Still, future work is needed to tease out these dynamics robustly.

Conclusion

Policy makers must consider the full range of public health consequences when regulating ride-hailing services and accordingly design appropriately nuanced interventions. While the observed drop in drunk driving is clearly beneficial, any causal association with increased binge drinking would be problematic. Two paths forward are thus evident. First, traditional health information campaigns have focused nearly exclusively on preventing people from getting behind the wheel. Drunk driving rates have been dropping for more than a decade, and this process may have been accelerated by ride-hailing. However, the evidence presented in this work indicates that messaging regarding alcohol consumption may need to be updated to address the growing problem of binge drinking. That is, while the information campaign against drunk driving should continue, our findings suggest that policy makers may instead wish to focus on the direct health effects of binge drinking. Second, it may be useful to partner with ride-hailing services to incentivize behavior through the platform. Beyond training drivers on how to appropriately manage and deal with intoxicated customers (eg, how to identify when they should be taken to hospital for medical care), opportunities exist to track excessive drinking at either the individual or local level (ie, behavioral surveillance) and incentivize more responsible behavior among citizens (ie, subsidized discounts for not drinking to excess). Critical to the success of these programs will be to ensure the platform is not imposing additional costs on binge drinkers, as this creates an incentive for them to return to driving under the influence.

We hope this work serves as a greater call to continue to explore the relationships between ride-hailing and public population health factors as well as the peer to peer sharing economy more broadly. Given that policy makers can only design effective interventions when presented with a full set of facts regarding the consequences of ride-hailing services, it is incumbent upon the research community to continue to provide rigorous insights, be they positive or negative, for practice or policy. Only in this way can we enable policy makers to implement steps that attenuate the negative aspects of ride-hailing and other digital platforms and amplify the positive aspects.

Acknowledgments

We gratefully acknowledge our anonymous reviewers' comments and appreciate their help in improving this article.

Authors' Contributions

GB and BNG originated the project and collected data. GB lead the statistical analysis with input from BNG and JM. All authors contributed equally to drafting the article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Uber entry schedule.
[DOCX File , 17 KB - jmir_v23i1e15402_app1.docx]

References

1. Burton R, Sheron N. No level of alcohol consumption improves health. Lancet 2018 Sep 22;392(10152):987-988 [FREE Full text] [doi: 10.1016/S0140-6736(18)31571-X] [Medline: 30146328]
2. Fairman BJ, Simons-Morton BG, Haynie DL, Liu D, Goldstein RB, Hingson RW, et al. State alcohol policies, taxes, and availability as predictors of adolescent binge drinking trajectories into early adulthood. Addiction 2019 Jul;114(7):1173-1182 [FREE Full text] [doi: 10.1111/add.14600] [Medline: 30830991]
3. Markowitz S, Grossman M. Alcohol regulation and domestic violence towards children. Contemp Econ Policy 1998;16(3):309-320. [doi: 10.1111/j.1465-7287.1998.tb00521.x]
4. Terza JV. Alcohol abuse and employment: a second look. J Appl Econ 2002 Jul;17(4):393-404. [doi: 10.1002/jae.671]
5. Monico N. Effects of alcohol on the body and mind. American Addiction Centers. 2020 Nov 03. URL: https://www.alcohol.org/effects/ [accessed 2021-01-06]
6. Hansen WB, Graham JW. Preventing alcohol, marijuana, and cigarette use among adolescents: peer pressure resistance training versus establishing conservative norms. Prev Med 1991 May;20(3):414-430. [doi: 10.1016/0091-7435(91)90039-7] [Medline: 1862062]
7. de Goeij MCM, Suhrcke M, Toffolutti V, van de Mheen D, Schoenmakers TM, Kunst AE. How economic crises affect alcohol consumption and alcohol-related health problems: a realist systematic review. Soc Sci Med 2015 Apr;131:131-146. [doi: 10.1016/j.socscimed.2015.02.025] [Medline: 25771482]
8. Jackson CK, Owens EG. One for the road: public transportation, alcohol consumption, and intoxicated driving. J Public Econ 2011 Feb;95(1-2):106-121. [doi: 10.1016/j.jpubeco.2010.09.010]
9. Barrios JM, Hochberg YV, Yi H. The cost of convenience: ridesharing and traffic fatalities. 2018. URL: http://yael-hochberg.com/assets/portfolio/BHY.pdf [accessed 2020-12-04]
10. Brazil N, Kirk DS. Uber and metropolitan traffic fatalities in the United States. Am J Epidemiol 2016 Aug 01;184(3):192-198. [doi: 10.1093/aje/kww062] [Medline: 27449416]

11. Dills AK, Mulholland SE. Ride-sharing, fatal crashes, and crime. Southern Econ J 2018 Jan 23;84(4):965-991. [doi: 10.1002/soej.12255]
12. Park J, Kim J, Lee B. Are uber really to blame for sexual assault? Evidence from New York city. Proceedings of the 18th Annual International Conference on Electronic Commerce: eCommerce in Smart connected World 2016:1. [doi: 10.1145/2971603.2971615]
13. Martin-Buck F. Driving safety: an empirical analysis of ridesharing's impact on drunk driving and alcohol-related crime. University of Texas at Austin Working Paper. 2017. URL: http://frankmartinbuck.com/Ridesharing%20and%20Alcohol-Related%20Crime%20by%20Frank%20Martin-Buck.pdf [accessed 2020-12-04]
14. Moskatel L, Slusky D. Did UberX reduce ambulance volume? Health Economics 2019 Jun 24;28(7):817-829. [doi: 10.1002/hec.3888]
15. Greenwood BN, Wattal S. Show me the way to go home: an empirical investigation of ride-sharing and alcohol related motor vehicle fatalities. MISQ 2017 Jan 1;41(1):163-187. [doi: 10.25300/MISQ/2017/41.1.08]
16. Burtch G, Carnahan S, Greenwood BN. Can you gig it? An empirical examination of the gig economy and entrepreneurial activity. Manag Sci 2018 Dec;64(12):5497-5520. [doi: 10.1287/mnsc.2017.2916]
17. Gong J, Greenwood B, Song Y. Uber might buy me a Mercedes Benz: an empirical investigation of the sharing economy and durable goods purchase. SSRN Journal 2017 May 19:1-47. [doi: 10.2139/ssrn.2971072]
18. Li Z, Hong Y, Zhang Z. An empirical analysis of the impacts of the sharing economy platforms on the US labor market. Proceedings of the 51st Hawaii International Conference on System Sciences 2018:1. [doi: 10.24251/hicss.2018.084]
19. Mejia J, Parker C. When transparency fails: bias and financial incentives in ridesharing platforms. SSRN Journal 2019:1. [doi: 10.2139/ssrn.3209274]
20. Greenwood B, Adjerid I, Angst C. How unbecoming of you: gender biases in perceptions of ridesharing performance. Proceedings of the 52nd Hawaii International Conference on System Sciences 2019:1. [doi: 10.24251/hicss.2019.789]
21. Cramer J, Krueger AB. Disruptive change in the taxi business: the case of Uber. Am Econ Rev 2016 May 01;106(5):177-182. [doi: 10.1257/aer.p20161002]
22. Wallsten S. The competitive effects of the sharing economy: how is Uber changing taxis? Technology Policy Institute. 2015. URL: https://techpolicyinstitute.org/wp-content/uploads/2017/07/Wallsten_The-Competitive-Effects-of-Uber.pdf [accessed 2020-12-04]
23. Babar Y, Burtch G. Examining the impact of ridehailing services on public transit use. Inf Syst Res 2020:1. [doi: 10.2139/ssrn.3042805]
24. Hall JD, Palsson C, Price J. Is Uber a substitute or complement for public transit? J Urban Econ 2018 Nov;108:36-50. [doi: 10.1016/j.jue.2018.09.003]
25. Rivara FP, Relyea-Chew A, Wang J, Riley S, Boisvert D, Gomez T. Drinking behaviors in young adults: the potential role of designated driver and safe ride home programs. Inj Prev 2007 Jun;13(3):168-172 [FREE Full text] [doi: 10.1136/ip.2006.015032] [Medline: 17567971]
26. Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD. 2010 national and state costs of excessive alcohol consumption. Am J Prev Med 2015 Nov;49(5):e73-e79. [doi: 10.1016/j.amepre.2015.05.031] [Medline: 26477807]
27. de la Merced MJ, Conger K. Uber I.P.O. Values Ride-Hailing Giant at $82.4 Billion. New York: New York Times; 2019 May 09. URL: https://www.nytimes.com/2019/05/09/technology/uber-ipo-stock-price.html [accessed 2021-01-15]
28. Bryan KA, Gans JS. A theory of multihoming in rideshare competition. J Econ Manage Strat 2019 Jan 08;28(1):89-96. [doi: 10.1111/jems.12306]
29. Angrist J. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press; 2008.
30. Autor D. Outsourcing at will: the contribution of unjust dismissal doctrine to the growth of employment outsourcing. J Labor Econ 2003 Jan;21(1):1-42. [doi: 10.1086/344122]
31. Gelman A, Stern H. The difference between "significant" and "not significant" is not itself statistically significant. Am Statistician 2006 Nov;60(4):328-331. [doi: 10.1198/000313006x152649]
32. Binge drinking fact sheet. Atlanta: Centers for Disease Control and Prevention; 2019. URL: https://www.cdc.gov/alcohol/fact-sheets/binge-drinking.htm [accessed 2019-03-29]
33. Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula MK, Marks JS. Binge drinking among US adults. JAMA 2003 Jan 1;289(1):70-75. [Medline: 12503979]
34. Abbey A, Ross LT, McDuffie D, McAuslan P. Alcohol and dating risk factors for sexual assault among college women. Psychol Women Q 2016 Jun 24;20(1):147-169. [doi: 10.1111/j.1471-6402.1996.tb00669.x]

Abbreviations

BRFSS: Behavioral Risk Factors Surveillance Survey
CDC: Centers for Disease Control and Prevention
MSA: Metropolitan Statistical Area

Edited by G Eysenbach, L Buis; submitted 08.07.19; peer-reviewed by D Slusky, C Quinn, PJ Schulz, M Parackal; comments to author 15.05.20; revised version received 21.09.20; accepted 25.10.20; published 27.01.21.

Please cite as:
Burtch G, Greenwood BN, McCullough JS
Ride-Hailing Services and Alcohol Consumption: Longitudinal Analysis
J Med Internet Res 2021;23(1):e15402
URL: http://www.jmir.org/2021/1/e15402/
doi:10.2196/15402
PMID:33502328

Original Paper

Patients' Perspectives About Factors Affecting Their Use of Electronic Personal Health Records in England: Qualitative Analysis

Alaa Abd-Alrazaq[1], BSc, MSc, PhD; Zeineb Safi[1], BSc, MSc; Bridgette M Bewick[2], BA, MA, PhD; Mowafa Househ[1], BCom, MEng, PhD; Peter H Gardner[3,4], BSc, MSc, PhD

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[2]Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom
[3]School of Pharmacy and Medical Sciences, University of Bradford, Bradford, United Kingdom
[4]Wolfson Centre for Applied Health Research, Bradford, United Kingdom

Corresponding Author:
Alaa Abd-Alrazaq, BSc, MSc, PhD
Division of Information and Computing Technology
College of Science and Engineering
Hamad Bin Khalifa University
LAS Building,
Education city
Doha, RD5256
Qatar
Phone: 974 55708549
Email: alaa_alzoubi88@yahoo.com

Abstract

Background: General practices (GPs) in England have recently introduced a nationwide electronic personal health record (ePHR) system called Patient Online or GP online services, which allows patients to view parts of their medical records, book appointments, and request prescription refills. Although this system is free of charge, its adoption rates are low. To improve patients' adoption and implementation success of the system, it is important to understand the factors affecting their use of the system.

Objective: The aim of this study is to explore patients' perspectives of factors affecting their use of ePHRs in England.

Methods: A cross-sectional survey was carried out between August 21 and September 26, 2017. A questionnaire was used in this survey to collect mainly quantitative data through closed-ended questions in addition to qualitative data through an open-ended question. A convenience sample was recruited in 4 GPs in West Yorkshire, England. Given that the quantitative data were analyzed in a previous study, we analyzed the qualitative data using thematic analysis.

Results: Of the 800 eligible patients invited to participate in the survey, 624 (78.0%) returned a fully completed questionnaire. Of those returned questionnaires, the open-ended question was answered by 136/624 (21.8%) participants. A total of 2 meta-themes emerged from participants' responses. The first meta-theme comprises 5 themes about why patients do not use Patient Online: concerns about using Patient Online, lack of awareness of Patient Online, challenges regarding internet and computers, perceived characteristics of nonusers, and preference for personal contact. The second meta-theme contains 1 theme about why patients use Patient Online: encouraging features of Patient Online.

Conclusions: The challenges and concerns that impede the use of Patient Online seem to be of greater importance than the facilitators that encourage its use. There are practical considerations that, if incorporated into the system, are likely to improve its adoption rate: Patient Online should be useful, easy to use, secure, and easy to access. Different channels should be used to increase the awareness of the system, and GPs should ease registration with the system and provide manuals, training sessions, and technical support. More research is needed to assess the effect of the new factors found in this study (eg, lack of trust, difficulty registering with Patient Online) and factors affecting the continuing use of the system.

(J Med Internet Res 2021;23(1):e17500) doi:10.2196/17500

KEYWORDS

electronic personal health records; tethered personal health records; patient portal; patient online; technology acceptance; technology adoption; qualitative research; mobile phone

Introduction

Background

Over the past 2 decades, there has been a rapid and widespread diffusion of electronic personal health records (ePHRs) in health care institutes [1]. The Markle Foundation defines ePHRs as web-based portals that enable users to access their medical records stored by their health care providers [2]. Other services can be added to ePHRs, such as booking appointments, requesting referrals, messaging health care providers, requesting medication refills, and educational materials [3,4]. Several benefits may be gained from using ePHRs, such as empowering patients [5,6], increasing their adherence to medication [7,8], improving their self-management [8,9], enhancing patient-provider relationships and communications [10,11], decreasing adverse events and allergic reactions [11,12], and avoiding duplicated tests [11,12].

General practices (GPs) in England started implementing ePHRs in 2003 when patients were enabled to access their full records through kiosks installed in some GPs. These kiosks allow patients to check their demographic information, consultations, test results, letters, and allergies [13].

In 2007, the National Health Service (NHS) offered patients in England access to their Summary Care Records (SCR) through HealthSpace [14-16]. HealthSpace is a secure web-based personal health record that has several functions: booking or canceling hospital appointments, recording and charting health indicators (eg, vital signs, weight, peak flow), calendar with email reminders, NHS address book, links to educational sources, secure messaging, and access to the SCR [15,17]. The SCR is a summary of key health information (allergies, adverse reactions, current medications, and main diagnoses) extracted from patient electronic medical records held by their general practitioners, and it is stored centrally and accessible by authorized NHS staff in urgent situations [14,16]. Because of the low adoption rate and technical issues, HealthSpace was shut down in December 2012 [18].

In 2015, the NHS implemented ePHRs under a program called Patient Online or GP online services, which enables users to book appointments, request prescription refills, and access coded information in their medical records such as demographics, medications, allergies, test results, problems list, immunizations, and medical and surgical procedures [19]. Currently, it is the largest ePHR in England, given that it has been implemented in more than 99% of GPs [19]. As the system is provided by different companies, it is called by different names such as Patient Access, Patient Services, The Waiting Room, and SystemOnline [19]. GP online services have been introduced in the United Kingdom at a time when funding for the NHS is under pressure. Given the context of austerity, individual practices have limited resources to support the rollout of GP online services.

Research Problem and Aim

Despite the potential benefits of ePHRs, their adoption rate in England was only 28% by the end of June 2019 [20]. Identifying the factors affecting patients' use of ePHRs is important to improve patients' adoption and the implementation success of ePHRs [21-25]. A systematic review of 97 studies found that factors affecting patients' use of ePHRs in England have not been examined, and there is a lack of qualitative studies (8%) in this topic [26]. Accordingly, this study aimed to explore patients' perspectives of factors affecting their use of ePHRs (Patient Online) in England.

Methods

Data Collection

A cross-sectional survey was conducted between August 21 and September 26, 2017. In this study, a self-administered questionnaire was used to collect quantitative data through closed-ended questions and qualitative data through an open-ended question (Multimedia Appendix 1). The qualitative data provide the focus of this study. Note that the findings from the quantitative analysis of the survey data were presented in a previous paper [27]. The survey gained health research authority approval before starting data collection (The Research Ethics Committee reference number: 17/SC/0323).

Sample

A convenience sample of patients was recruited from 4 GPs in West Yorkshire, England. Patients were eligible to participate if they (1) lived in England and were registered at 1 of the 4 GPs, (2) were aged 18 years or older, and (3) had not used Patient Online before (nonusers).

Analysis

The qualitative data were analyzed using thematic analysis. Given the exploratory nature of this study, an inductive approach was used to generate themes directly from the data [28]. The analysis was performed following the steps proposed by Braun and Clarke [29]: (1) familiarizing with the data through scrutinizing and rescrutinizing the transcript; (2) coding data systematically; (3) generating subthemes and themes from codes; (4) checking the fit of those themes and subthemes to the original utterances and drawing an initial thematic map; (5) refining and regrouping some inappropriate codes and generating meta-themes from the themes for more granular grouping; and finally, (6) defining and naming subthemes, themes, and meta-themes. We followed the guidelines of Braun and Clarke, as these are considered the most systematic guide for conducting thematic analysis to date [30,31]. The analysis was carried out by the first author (AA), and the validity of codes and themes was checked by another author (BB). AA and BB discussed codes and themes. Where AA and BB had differing views on the code labels and/or thematic content, these discrepancies were resolved through discussion. In all cases, agreement was reached between AA and BB. Microsoft Excel was used to manage the analysis process.

Results

Collected Data

Out of the 800 eligible patients invited to participate in the survey, 624 (78%) participants completed the questionnaire. Of those participants, 136 (21.8%) answered the open-ended question. The 136 comments contained 221 utterances. A comment refers to the whole text written by a participant as a reply to our question, whereas an utterance refers to a part of the comment that has one idea or thought. In total, 3 of the 221 utterances were excluded because 2 utterances were illegible and the meaning of 1 utterance was not discernible. The final number of utterances included in the thematic analyses was 218. The excluded utterances were all part of longer comments, and for that reason, the final number of comments remained 136. Subsection 3.1 summarizes the characteristics of the respondents, and Subsection 3.2 presents the findings of the thematic analysis.

Participants’ Characteristics

Table 1 summarizes the characteristics of the participants who answered the open-ended question and those who did not. Those who responded to the question had a mean age of 43.7 years (SD 18.3). More female participants answered the question than male participants (80/136, 58.8% were females). The majority of the respondents had a White ethnicity (107/136, 78.7%), had an income of less than US $40,000 per year (95/136, 69.8%), and had access to the internet (112/136, 82.4%). In terms of education, 39.7% (54/136) of the respondents had a bachelor’s degree or higher.

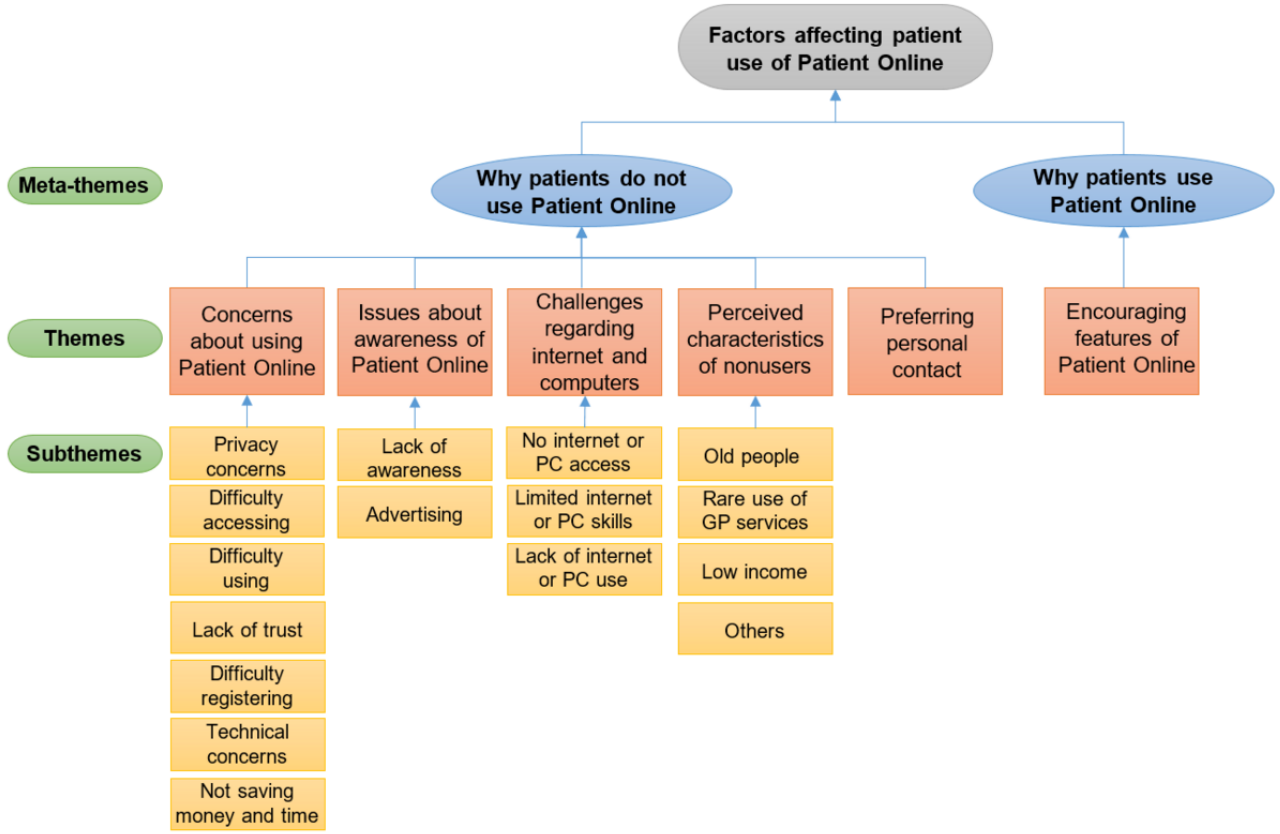
Table 1. Characteristics of respondents (n=136).

Characteristics	Value, n (%)
Age (years), mean (SD)	43.7 (18.3)
18-24	19 (14.0)
25-34	35 (25.7)
35-44	23 (16.9)
45-54	20 (14.7)
55-64	17 (12.5)
65-74	12 (8.8)
≥75	10 (7.4)
Sex	
Male	56 (41.2)
Female	80 (58.8)
Ethnicity	
White	107 (78.7)
Asian	14 (10.3)
Black	6 (4.4)
Mixed	7 (5.1)
Others	2 (1.5)
Income (US $)	
<20,000	55 (40.4)
20,000-29,999	24 (17.6)
30,000-39,999	16 (11.8)
40,000-49,999	9 (6.6)
50,000-59,999	6 (4.4)
60,000 or more	5 (3.8)
Prefer not to say	21 (15.4)
Education	
Up to secondary school	13 (9.6)
Secondary school	31 (22.8)
College/Diploma	38 (27.9)
Bachelor’s degree	38 (27.9)
Master’s degree	10 (7.4)
Doctoral degree	6 (4.4)
Internet access	
Yes	112 (82.4)
No	24 (17.6)

Findings of Thematic Analysis

In total, 2 meta-themes were generated as a result of the thematic analysis. The first meta-theme consists of 5 themes and relates to utterances explaining why patients do not use Patient Online (Figure 1). The second meta-theme pertains to utterances about why patients use Patient Online, and it contains 1 theme: encouraging features of Patient Online. The following sections contain more details about all 6 themes.

Figure 1. Thematic map. GP: general practice.



Theme 1: Concerns About Using Patient Online

The first theme, concerns about using Patient Online, is made up of 7 subthemes: (1) concerns about privacy and security, (2) difficulty accessing Patient Online, (3) difficulty using Patient Online, (4) lack of trust in Patient Online, (5) difficulty registering, (6) technical concerns, and (7) the inability of Patient Online to save money and time (Figure 1).

The security and privacy of Patient Online was a major concern for respondents. Their concerns were attributed to the recent NHS hack attacks, worries that their data will be accessed by third parties, and uncertainty about the security measures of Patient Online:

> *I believe that Patient Online has/ will have too many privacy issues, look what happened when the NHS was hacked.* [Participant #9]
>
> *Only concern is confidentiality of System One as I am aware CIA [Central Intelligence Agency] are now using the system.* [Participant #30]

The second subtheme shows that difficulty accessing (logging in) the system can be a barrier to its use. The main reasons given for difficulty accessing Patient Online were the inability to find its URL link and forgetting passwords and log-in details:

> *I tried to use the system but I can never find the correct link...* [Participant #120]
>
> *...I always forget my password.* [Participant #35]

The third subtheme was generated from comments about difficulty using Patient Online. Although the previous subtheme reflects patients' concerns about logging on to Patient Online, this subtheme represents their worries about using the system after logging in to it (ie, ordering prescriptions, managing appointments, checking their records). According to some respondents, these concerns are exacerbated when nobody can help in using the system:

> *I don't know if this would be easy to use.* [Participant #5]
>
> *If people experience a difficulty and do not know where to find help, or who to ask, they may give up trying.* [Participant #49]

The fourth subtheme indicates that some patients did not trust Patient Online to do what they want it to do. They doubted that an appointment would actually be booked for them if they booked via Patient Online:

> *...I don't trust the service.* [Participant #9]
>
> *...I am not sure I would entirely trust it...* [Participant #123]

Concerns with difficulty registering with Patient Online were brought together to form the fifth subtheme. These concerns were attributed to the fact that they have to visit the practice in person with their ID to be able to register to use the system. To ease the registration process, a respondent suggested that the registration to Patient Online becomes part of the patient registration in practice:

> *You also have to make a trip to the surgery with ID to be able to use the service.* [Participant #28]

> *I think more effort should be made to encourage patients to sign up for this, and the process should be more streamlined-perhaps done as a matter of course when registering.* [Participant #7]

The sixth subtheme encompasses utterances that show concern regarding the technical difficulties of Patient Online. Technical issues here refer to technical errors that people believe they will face when using Patient Online:

> *Technology goes wrong and does not tell you why.* [Participant #58]

The last subtheme brought together utterances from some respondents who were worried about the inability of Patient Online to save money and time. This is reflected in the utterances of the seventh subtheme, that is, respondents, especially those who live near the practice doubted that using Patient Online saves money and time:

> *In my experience many of these things do not end up saving people's time and money. So I don't think I'll be using this except infrequently.* [Participant #38]
>
> *It would not save travel costs because I live next to it.* [Participant #85]

Theme 2: Issues About Awareness of Patient Online

The utterances in this theme suggest that if respondents had more knowledge or awareness about Patient Online, they would use it. This theme consists of 2 subthemes: lack of awareness of Patient Online and advertising about Patient Online. In the first subtheme, the respondents stated that the lack of knowledge about the system's presence, what it is about, how to use, and how to access it was the main reason for not using it:

> *To be honest, I've never heard of Patient Online before and that may be why people haven't used it.* [Participant #88]
>
> *Not been shown what it is about and how to use it.* [Participant #80]

In the second subtheme, several respondents attributed their lack of awareness of Patient Online to the lack of advertisement about it. For this reason, they acknowledged the essential role of the publicity of Patient Online in increasing people's awareness of it:

> *It is not openly advertised in the surgery that Patient Online is available.* [Participant #28]
>
> *...may not be enough advertisement.* [Participant #62]

Theme 3: Challenges Regarding Internet and Computers

The third theme refers to issues regarding prerequisites for using Patient Online (ie, a computer and internet access). Respondents identified 3 challenges regarding the internet and computers, which form the 3 subthemes within this theme. The first challenge is the lack of internet or computer access. Many respondents attributed the nonuse of Patient Online to not having internet or computer access:

> *Those who don't have access to the internet may not use it.* [Participant #57]

Although many respondents have access to the internet and computers, they have limited skills in using them, and this is the second challenge:

> *I can't use a computer so I can't use Patient Online.* [Participant #2]

The third challenge is the lack of use of internet or computers. This subtheme indicates that some users may have access to computers and internet and the required skills but do not frequently use them:

> *I do not use computers of any kind.* [Participant #75]

Theme 4: Perceived Characteristics of Nonusers

The fourth theme was generated from utterances about who is less likely to use Patient Online. The 3 main characteristics of nonusers were related to age, use of GP services, and income. These characteristics formed 3 subthemes, in addition to an extra subtheme that encompasses infrequently reported characteristics.

Age was the most commonly reported characteristic of nonusers. Respondents suspected the ability of older people to use Patient Online for different reasons: lack of computer and internet skills, lack of internet access, lack of awareness of how to use the system, lack of confidence in using it, lack of technology use, and their preference for face-to-face contact:

> *Elderly people may have no understanding or knowledge of how to use a computer or the internet.* [Participant #69]
>
> *Older people may not use it as they don't have access to internet or know how to use services.* [Participant #116]

In the second subtheme, respondents attributed the nonuse of the system to lack of use of GP services in general, such as consultations and medications:

> *I've never used it as it's rare that I attend the surgery and I'm not on any medication.* [Participant #132]

Low income formed the third subtheme. Respondents stated that people need enough income to have internet access or get training to be able to use computers and the internet:

> *I do not have enough income/benefits...* [Participant #20]

The last subtheme encompasses characteristics of other people who are more likely to be nonusers of Patient Online and those who live near the practice, illiterate people, people who cannot read in English, and people who forget to use Patient Online:

> *I would use Patient Online more often if I lived further away from the surgery.* [Participant #15]
>
> *I am not good at reading or spelling so online would not be good for me.* [Participant #70]

Theme 5: Preferring Personal Contact

Preferring personal contact was identified as the main justification for not using Patient Online. Patients prefer personal contact because they think it is more reliable, easier, provides an instant reply, and is important in urgent conditions:

> *It is more reliable to speak to someone directly about their medical records rather than using online.* [Participant #29]
>
> *Picking up the phone and speaking to someone is easier.* [Participant #135]

Theme 6: Encouraging Features of Patient Online

Within this theme, respondents identified features of Patient Online that may encourage them to use the system. One of the main features of Patient Online is that it is useful for different people, such as students, people with mobility needs, people who cannot reach the practice, and busy people:

> *I feel that it would be particularly useful for students.* [Participant #63]

Another feature mentioned by respondents is ease of access. Some respondents thought that Patient Online could be more accessible if it was a mobile app. It is noteworthy that mobile apps were not available for accessing GP online services at the time of data collection. Later, patients have been enabled to access GP online services via a mobile app called NHS App [32]:

> *A mobile application would be more accessible...* [Participant #95]

Respondents reported other features of Patient Online, which may encourage people to use it, namely, secure, quick, user-friendly interface, convenient, and less stressful:

> *If it is secure and fast then people will use it, I suppose.* [Participant #68]
>
> *If the interface is not user-friendly people might not be encouraged to use Patient Online.* [Participant #82]

Discussion

Principal Findings

The aim of this study is to explore why patients in England choose to use ePHRs. Participants identified one leading cause that encouraged them to use Patient Online, which relates to its features being useful, easy to access, secure, quick, user-friendly interface, convenient, and less stressful. However, patients identified many reasons for not using Patient Online, which were categorized into 5 themes: concerns about using Patient Online, lack of awareness of Patient Online, challenges regarding internet and computers, perceived characteristics of nonusers, and preference for personal contact.

In the first theme, concerns about using Patient Online, the most prominent reason for not using Patient Online was privacy and security concerns. This may be attributed to the fact that ePHRs typically contain personal and sensitive information, and patients have previously been shown to be concerned about the accessibility of these data [33]. The hack attacks that happened to the NHS 4 months before data collection may have exacerbated these concerns in this sample. This finding is consistent with the results of the quantitative data in the original study [27], where perceived privacy and security significantly affected patients' intention to use Patient Online. This factor was also found in other quantitative studies [33-36] and qualitative studies [37-43].

Participants also raised their concerns about difficulty logging on to Patient Online because of losing its URL and forgetting passwords and log-in details. This issue posed a challenge for patients because they were given new complex passwords and usernames to access Patient Online. Although passwords can be changed through the system, usernames are fixed. This effect of difficulty accessing the system has also been demonstrated in several studies [41,44,45].

Other worries were reported about difficulty using Patient Online, especially when there is no one to help. This may be attributed to the fact that patients need adequate computer and internet skills to use Patient Online. They may also need to access it without any help from others to protect their privacy. This factor was also found in quantitative analysis in the original study [27], where effort expectancy (ie, ease of use) and behavioral intentions were significantly associated. Furthermore, numerous quantitative and qualitative studies have shown similar findings regarding this factor [37,38,40,41,45-47].

Participants expressed their concerns about the difficulty they experienced registering with Patient Online. Indeed, it could be argued that the process of registration with Patient Online is less flexible than several systems (eg, MyChart, PatientSite, My Health Manager, My Health at Vanderbilt), where patients can register with the system using email, websites, or phone and with no need to visit the practice. To the best of our knowledge, this factor was not found in previous studies. This may be because of the ease of registration with other systems.

The inability of Patient Online to save money and time was a concern for some participants, especially those who live near the practice. This concern may have made patients feel that Patient Online is not useful for them. Thus, this factor is related to perceived usefulness, which was the most influential factor according to the quantitative analysis in the original study [27]. The effect of this factor was also demonstrated by quantitative studies [46-49] and other qualitative studies [37,40,41,45,50].

Finally, 2 further concerns in this group were raised by participants, a lack of trust in Patient Online to book appointments or request medication refills and the technical issues that some patients reported when using Patient Online. To the best of our knowledge, neither of these factors have been reported in previous studies.

In the second theme, lack of awareness of different aspects of Patient Online was an influential factor in not using the system. Lack of advertising about Patient Online was the main reason for this lack of awareness. Although 3 of the 4 GPs had advertisements about Patient Online visibly displayed on screens or brochures in the waiting room during the study, some patients still reported a lack of awareness of the system. This factor was in line with the findings of previous quantitative studies [51,52] and qualitative studies [37,41,45,53].

With regard to the third theme, 3 challenges related to computers and the internet were identified. The first is the lack of internet or computer access. This factor was represented by the construct *facilitating condition* in the quantitative analysis, and it was found to significantly affect the actual use of Patient Online [27]. Furthermore, previous studies have shown a significant deleterious effect of a lack of internet [54-59] and computer access [38,46,52,57].

The limited skills in using the internet or computers was the second challenge in this group. This challenge may have produced reports that patients found Patient Online difficult to use. Hence, this factor is related to perceived ease of use (ie, effort expectancy), which was the most influential factor according to the quantitative analysis in the original study [19]. Numerous studies have supported this effect of computer literacy [38,40,41,53,60] and internet literacy [61,62].

The last challenge was the lack of using internet or computers. This challenge may also be related to perceived ease of use, as those who rarely use computers and the internet may perceive the system difficult to use. Several previous studies showed similar findings regarding the effect of lack of computer use [43,46,62] and internet use on the adoption of ePHRs [39,43,54,63-65].

Regarding the fourth theme, participants determined the following characteristics of nonusers of Patient Online, which were consistent with findings of previous studies: older people [61,66-69], who rarely use GP services [55,66,68,70,71], who have low income [46,52,72,73], who live near the practice [70], and who have lower literacy levels [46,52,72,73].

In the last theme, participants justified their nonuse of the system by indicating their preference for personal contact with their GP. This was attributed to the perceived advantages of personal contact over the system. This factor was found in other studies [40,51,52].

Strengths

This study enabled us to explore new factors that were not examined by the quantitative part of the study (eg, lack of awareness) and previous studies (eg, lack of trust). Furthermore, this study allowed us, to some extent, to support and explain some relationships proposed in the quantitative study (eg, performance expectancy, perceived privacy, security).

To the best of our knowledge, this study had the largest sample size in comparison with all qualitative studies on this topic. This allowed us to explore a wide range of patients' perspectives on the adoption of ePHRs.

Limitations

This study collected data from 4 GPs implementing the same ePHR (ie, SystemOnline), which may limit the generalizability of this study to other practices implementing other ePHRs (ie, Patient Access, Patient Services, The Waiting Room, Engage Consult, and Evergreen Life/i-Patient). However, it should be noted that all these systems provide the same services to the patients (ie, booking appointments, requesting prescription refills, and viewing health records), and no participant had used any of them before. As a result, the participants in this study were unlikely to have made comparisons between the different systems.

Although the qualitative data collected by an open-ended question helped in exploring factors affecting patients' use of Patient Online, such data may not be equivalent to qualitative data collected by interviews or focus groups. Thus, we could not deeply understand the adoption process of Patient Online. However, this qualitative analysis did not aim to understand in depth the phenomenon of interest; rather, it aimed only to help in identifying other factors not included in the model and explaining the findings of the quantitative study. As answering the open-ended question was voluntary, there may be an element of self-selection.

As the open-ended question was put after closed-ended questions, participants' answers to the open-ended question may be influenced by this order. This order was based on researchers' recommendations that questionnaires should start with the most interesting and easy-to-answer questions, and open-ended and demographic questions should be presented at the end of the questionnaire [74-76].

Practical Implications

We believe that adoption of GP online services will significantly increase in the future, given that many factors identified in this study will be automatically and considerably mitigated by time. Specifically, the proportion of patients who are more comfortable with the use of computers, smartphones, electronic systems, and the internet will increase in the future given their increased spread over the world. Thus, these services may be desired and expected by patients. However, developers, marketers, and GPs still play a crucial role in increasing the adoption of GP online services.

During system development, patients should be involved in the process to identify the features that make the system useful and easy to use. Some participants pointed out that the system will be useful when it allows them to book walk-in appointments, communicate with their doctors, and select the required doctor. As Patient Online currently enables patients to choose the required doctor, developers should consider adding these services, which are provided by many ePHRs (eg, MyChart, MyHealtheVet, Patient Gateway) [46,66,77]. Furthermore, users of such systems should be informed and reassured about the different security measures that are in place (eg, strong firewalls, encouragement to use complex and long passwords), and it should be made clear that the provision of GP online services is strictly controlled by legislation to safeguard personal data. To ease logging on to the system, developers should develop a system that allows patients to access it through their fingerprints or face recognitions, instead of using complex usernames and passwords. It is noteworthy that the NHS App, which has been recently developed, is the only system that enables patients to access GP online services using fingerprints or face recognitions [32].

To increase the awareness of the system, its functionality, and its benefits, marketers should improve their publicity through different channels, such as public media (eg, television, radio, newspapers, magazines), social media (eg, Facebook, Twitter, YouTube), emails, mails, automated messages on the practices' telephone system, and advertisements in general public areas (eg, shopping centers, health care settings, highway streets, universities). Face-to-face communication is considered as one of the most effective channels in marketing to persuade potential adopters to adopt an innovation [78,79]. Thus, all staff in practice (eg, physicians, nurses, receptionists) should offer the system to patients during their visits. GP staff may not be keen on publicizing online services because of a lack of incentives and time. Therefore, consideration should be given to providing incentives and resources for GPs to increase patients' awareness of GP online services.

Although patients have been recently enabled to sign up in the GP online services without visiting their surgeries through only the NHS App [80], they still need to visit their surgeries in person to register to use GP online services provided by other systems (eg, SystemOnline, Patient Access). To ease signing up in these systems, GPs should allow patients to register on web or through phone and make the signing up procedure a part of patient registration in the practice. GPs may enhance patients' perceptions of usefulness, ease of use of the system, and their trust in it by helping them in using a beta version of the system through a computer in a waiting room. GPs should provide online assistance, technical support, manuals, and training to allow patients to solve any technical issues that face them when using the system, thereby decreasing their technical concerns. GPs should collaborate with other parties (eg, Patient Online providers and government bodies) to provide computers and/or internet access at affordable prices for those who do not have them and cannot afford them. Given that many UK GPs report being overstretched and limited funding has been provided to support the rollout of GP online services, consideration should be given to providing incentive programs (eg, Meaningful Use policy as issued by the US government). Incentive programs could be used to encourage GPs to publicize their online services and encourage patients to use them.

Recommendations for Future Research

As this study could not provide a deep understanding of the adoption process of Patient Online, a deeper understanding of the adoption of online services could be gained through further qualitative work using interviews or focus groups. Several factors were revealed in this analysis but were not part of the conceptual model in the quantitative study, namely, awareness of Patient Online, lack of trust in the system, difficulty registering, disability, lack of use of GP services, and distance to the GPs. Future studies should consider adding these factors to the model and quantitatively examine them. Finally, more research is needed to identify the factors affecting the continuing use, as long-term viability and eventual success of information technology count on its continuing use more than initial use [81-83].

Conclusions

This research explored patients' perspectives regarding factors influencing their use of Patient Online. We found about 20 factors grouped into 6 themes. The findings of this study supported the findings of the quantitative study (eg, performance expectancy, effort expectancy, perceived privacy). This study found new factors that were not examined by the quantitative part of the study (eg, lack of awareness) and previous studies (eg, lack of trust).

The challenges and concerns that impede the use of Patient Online seem to be greater than the facilitators that encourage its use. To foster use, several practical implications were suggested: Patient Online should be useful, easy to use, secure, and easy to access; different channels should be used to increase the awareness of the system; and GPs should ease registration with the system and provide manuals, training sessions, and technical support. More research is needed to quantitatively assess the effect of the new factors found in this study (eg, lack of trust, difficulty registering with Patient Online) and factors affecting continuing use of the system.

Acknowledgments

The authors would like to thank Dr Hamish Fraser for his help in the initial stages of the study, especially for his contribution to the development of the study design and setting up of links with practices. The authors would also like to thank Tracey Farragher for her help in assessing the nonresponse bias. The publication of this study was funded by the Qatar National Library.

Authors' Contributions

AA and PG developed the protocol. AA collected the data. AA analyzed the data under BB's guidance and supervision. AA and ZS drafted the manuscript, and it was revised critically for important intellectual content by all authors. All authors approved the manuscript for publication and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 70 KB - jmir_v23i1e17500_app1.docx]

References

1. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
2. The Personal Health Working Group: Final Report. Markle Foundation. 2003. URL: https://www.markle.org/sites/default/files/final_phwg_report1.pdf [accessed 2020-12-21]
3. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to electronic health records. Med Care 2013;51:S52-S56. [doi: 10.1097/mlr.0b013e31827808db]
4. Pai H, Lau F, Barnett J, Jones S. Meeting the health information needs of prostate cancer patients using personal health records. Curr Oncol 2013 Dec;20(6):e561-e569 [FREE Full text] [doi: 10.3747/co.20.1584] [Medline: 24311957]
5. Ertmer A, Uckert F. User acceptance of and satisfaction with a personal electronic health record. Stud Health Technol Inform 2005;116:861-866. [Medline: 16160366]
6. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of innovators adopting a national personal health record in Portugal: cross-sectional study. JMIR Med Inform 2017 Oct 11;5(4):e37 [FREE Full text] [doi: 10.2196/medinform.7887] [Medline: 29021125]
7. Rice R. Healthcare Leaders' Lived Experiences Regarding the Implementation of Electronic Personal Health Records. New York, USA: Walden University; 2014.
8. Vermeir P, Degroote S, Vandijck D, Van Tiggelen H, Peleman R, Verhaeghe R, et al. The patient perspective on the effects of medical record accessibility: a systematic review. Acta Clin Belg 2017 Jun;72(3):186-194. [doi: 10.1080/17843286.2016.1275375] [Medline: 28056665]
9. Smith AB, Odlum M, Sikka M, Bakken S, Kanter T. Patient perceptions of pre-implementation of personal health records (PHRS): a qualitative study of people living with HIV in New York city. J HIV/AIDS Soc Serv 2012 Oct;11(4):406-423. [doi: 10.1080/15381501.2012.735166]
10. Ochoa IA, Kitayama K, Uijtdehaage S, Vermillion M, Eaton M, Carpio F, et al. Patient and provider perspectives on the potential value and use of a bilingual online patient portal in a Spanish-speaking safety-net population. J Am Med Inform Assoc 2017 Nov 1;24(6):1160-1164 [FREE Full text] [doi: 10.1093/jamia/ocx040] [Medline: 28460130]
11. Alyami M. Removing Barriers in Using Personal Health Record Systems. In: 15th International Conference on Computer and Information Science. 2016 Presented at: ICIS'16; 25 August 2016; Okayama, Japan. [doi: 10.1109/icis.2016.7550810]
12. Morton A. Examining Acceptance of an Integrated Personal Health Record (PHR). New York, USA: University of Maryland, Baltimore; 2011.
13. Fisher B, Fitton R, Poirier C, Stables D. Patient record access--the time has come!. Br J Gen Pract 2007 Jun;57(539):507-511 [FREE Full text] [Medline: 17550687]
14. Kahn JS, Aulakh V, Bosworth A. What it takes: characteristics of the ideal personal health record. Health Aff (Millwood) 2009;28(2):369-376. [doi: 10.1377/hlthaff.28.2.369] [Medline: 19275992]
15. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. Br Med J 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
16. Greenhalgh T, Wood GW, Bratan T, Stramer K, Hinder S. Patients' attitudes to the summary care record and HealthSpace: qualitative study. Br Med J 2008 Jun 7;336(7656):1290-1295 [FREE Full text] [doi: 10.1136/bmj.a114] [Medline: 18511764]
17. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. Br Med J 2010 Nov 16;341:c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
18. de Lusignan S, Ross P, Shifrin M, Hercigonja-Szekeres M, Seroussi B. A comparison of approaches to providing patients access to summary care records across old and new europe: an exploration of facilitators and barriers to implementation. Stud Health Technol Inform 2013;192:397-401. [Medline: 23920584]
19. Start Using GP Online Services. NHS. 2016. URL: https://www.nhs.uk/using-the-nhs/nhs-services/gps/gp-online-services/ [accessed 2020-12-21]
20. Patient Online Management Information (POMI). TinyURL. 2019. URL: https://tinyurl.com/ybptnhsu [accessed 2020-12-21]
21. Fung V, Ortiz E, Huang J, Fireman B, Miller R, Selby JV, et al. Early experiences with e-health services (1999-2002): promise, reality, and implications. Med Care 2006 May;44(5):491-496. [doi: 10.1097/01.mlr.0000207917.28988.dd] [Medline: 16641669]
22. Huygens MW, Vermeulen J, Friele RD, van Schayck OC, de Jong JD, de Witte LP. Internet services for communicating with the general practice: barely noticed and used by patients. Interact J Med Res 2015 Nov 24;4(4):e21 [FREE Full text] [doi: 10.2196/ijmr.4245] [Medline: 26601596]
23. Najaftorkaman M, Ghapanchi A, Talaei-Khoei A. Analysis of Research in Adoption of Person-Centred Healthcare Systems. Open Repository. 2014. URL: https://openrepository.aut.ac.nz/handle/10292/8104 [accessed 2020-12-30]
24. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
25. Logue MD, Effken JA. Modeling factors that influence personal health records adoption. Comput Inf Nurs 2012:1. [doi: 10.1097/nxn.0b013e3182510717]

26. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
27. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in England: cross-sectional study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
28. Boyatzis R. Thematic Analysis and Code Development. Thousand Oaks, CA: Sage Publications; 1998.
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
30. Howitt D. Introduction to Qualitative Methods in Psychology. London, UK: Pearson; 2016.
31. Wilson S, MacLean R. Research Methods and Data Analysis for Psychology. New York, USA: McGraw-Hill Education; 2011.
32. The NHS App. 2020. URL: https://www.nhs.uk/using-the-nhs/nhs-services/the-nhs-app/ [accessed 2020-12-21]
33. Daglish D. Electronic Personal Health Records: A Matter of Trust. McMaster University. 2013. URL: https://macsphere.mcmaster.ca/handle/11375/13325 [accessed 2020-12-30]
34. Rao M. Factors Affecting Health Care Technology Use in Baby Boomers: A Quantitative Study. New York, USA: Northcentral University; 2014.
35. Whetstone M, Goldsmith R. Factors influencing intention to use personal health records. Intl J of Pharm Health Mrkt 2009 Apr 3;3(1):8-25. [doi: 10.1108/17506120910948485]
36. Ozok AA, Wu H, Gurses AP. Exploring patients’ use intention of personal health record systems: implications for design. Int J Hum-Comput Interact 2017 Jan 6;33(4):265-279. [doi: 10.1080/10447318.2016.1277637]
37. Dontje K, Corser WD, Holzman G. Understanding patient perceptions of the electronic personal health record. J Nurs Pract 2014 Nov;10(10):824-828. [doi: 10.1016/j.nurpra.2014.09.009]
38. Mishuris RG, Stewart M, Fix GM, Marcello T, McInnes DK, Hogan TP, et al. Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study. Health Expect 2015 Dec;18(6):2296-2305 [FREE Full text] [doi: 10.1111/hex.12199] [Medline: 24816246]
39. Patel VN, Dhopeshwarkar RV, Edwards A, Barrón Y, Sparenborg J, Kaushal R. Consumer support for health information exchange and personal health records: a regional health information organization survey. J Med Syst 2012 Jun;36(3):1043-1052. [doi: 10.1007/s10916-010-9566-0] [Medline: 20703633]
40. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 3;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
41. Turner A, Osterhage K, Joe J, Hartzler A, Lin L, Demiris G. Use of patient portals: personal health information management in older adults. Stud Health Technol Inform 2015;216:978. [Medline: 26262280]
42. Damron BL, Kelly LS. Short-term exposure of laying hens to high dietary sodium chloride levels. Poult Sci 1987 May;66(5):825-828 [FREE Full text] [doi: 10.3382/ps.0660825] [Medline: 3628162]
43. Nguyen Q, Bartlett G, Rodriguez C, Tellier P. Young adults on the perceived benefits and expected use of personal health records: a qualitative descriptive study. J Innov Health Inform 2016 Apr 18;23(1):171 [FREE Full text] [doi: 10.14236/jhi.v23i1.171] [Medline: 27348486]
44. Ronda MC, Dijkhorst-Oei L, Rutten GE. Patients' experiences with and attitudes towards a diabetes patient web portal. PLoS One 2015;10(6):e0129403 [FREE Full text] [doi: 10.1371/journal.pone.0129403] [Medline: 26086272]
45. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health 2007 Oct;13(5):509-517. [doi: 10.1089/tmj.2006.0089] [Medline: 17999613]
46. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012 Nov 5;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]
47. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 2;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
48. Klein R. An empirical examination of patient-physician portal acceptance. Eur J Inf Syst 2017 Dec 19;16(6):751-760. [doi: 10.1057/palgrave.ejis.3000719]
49. Lazard A, Watkins I, Mackert M, Xie B, Stephens K, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):e157-e161 [FREE Full text] [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
50. Zickmund SL, Hess R, Bryce CL, McTigue K, Olshansky E, Fitzgerald K, et al. Interest in the use of computerized patient portals: role of the provider-patient relationship. J Gen Intern Med 2008 Jan;23(Suppl 1):20-26 [FREE Full text] [doi: 10.1007/s11606-007-0273-6] [Medline: 18095039]

51. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
52. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care 2012;50(5):434-440. [doi: 10.1097/mlr.0b013e318249d81b]
53. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013 Jul 3;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
54. Nambisan P. Factors that impact Patient Web Portal Readiness (PWPR) among the underserved. Int J Med Inform 2017 Jun;102:62-70. [doi: 10.1016/j.ijmedinf.2017.03.004] [Medline: 28495349]
55. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
56. Perzynski A, Roach M, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc 2017 Sep 1;24(5):927-932. [doi: 10.1093/jamia/ocx020] [Medline: 28371853]
57. Ronda MC, Dijkhorst-Oei L, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
58. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Assoc 2009 Sep 1;16(5):683-689. [doi: 10.1197/jamia.m3169]
59. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The association of patient factors, digital access, and online behavior on sustained patient portal use: a prospective cohort of enrolled users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]
60. Gu Y, Day K. Propensity of people with long-term conditions to use personal health records. Stud Health Technol Inf 2013;188:46-51.
61. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 4;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
62. Butler J, Carter M, Hayden C, Gibson B, Weir C, Snow L. Understanding adoption of a personal health record in rural health care clinics: revealing barriers and facilitators of adoption including attributions about potential patient portal users and self-reported characteristics of early adopting users. In: AMIA Annual Symposium Proceedings. 2013 Presented at: AMIA'13; May 6-9, 2013; New York, USA p. 152-161.
63. Arcury TA, Quandt SA, Sandberg JC, Miller DP, Latulipe C, Leng X, et al. Patient portal utilization among ethnically diverse low income older adults: observational study. JMIR Med Inform 2017 Nov 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
64. Kim S, Kim HR, Bae J, Kim Y. The consumers' perceptions and requirements for personal health records in Korea. J Korean Soc Med Inform 2009;15(3):273. [doi: 10.4258/jksmi.2009.15.3.273]
65. Sanders MR, Winters P, Fortuna RJ, Mendoza M, Berliant M, Clark L, et al. Internet access and patient portal readiness among patients in a group of inner-city safety-net practices. J Ambulat Care Manag 2013;36(3):251-259. [doi: 10.1097/jac.0b013e31829702f9]
66. Ancker JS, Nosal S, Hauser D, Way C, Calman N. Access policy and the digital divide in patient access to medical records. Health Pol Technol 2017 Mar;6(1):3-11. [doi: 10.1016/j.hlpt.2016.11.004]
67. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
68. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. CJASN 2015 Oct 22;10(11):2013-2022. [doi: 10.2215/cjn.01640215]
69. Leveille SG, Mejilla R, Ngo L, Fossa A, Elmore JG, Darer J, et al. Do patients who access clinical information on patient internet portals have more primary care visits? Medical Care 2016;54(1):17-23. [doi: 10.1097/mlr.0000000000000442]
70. Garrido T, Kanter M, Meng D, Turley M, Wang J, Sue V, et al. Race/ethnicity, personal health record access, and quality of care. Am J Manag Care 2015 Feb 1;21(2):e103-e113 [FREE Full text] [Medline: 25880485]
71. Martinez M, Baum A, Saldaño A, Gómez A, Luna D, de QF. Predictive variables of the use of personal health record: the Hospital Italiano de Buenos Aires study. Stud Health Technol Inf 2013:1171.
72. Ancker JS, Osorio SN, Cheriff A, Cole CL, Silver M, Kaushal R. Patient activation and use of an electronic patient portal. Inform Health Soc Care 2015;40(3):254-266. [doi: 10.3109/17538157.2014.908200] [Medline: 24786648]
73. Cahill J, Lin L, LoBiondo-Wood G, Armstrong T, Acquaye A, Vera-Bolanos E, et al. Personal health records, symptoms, uncertainty, and mood in brain tumor patients. Neurooncol Pract 2014 Jun;1(2):64-70 [FREE Full text] [doi: 10.1093/nop/npu005] [Medline: 26034618]
74. Bowling A. Research Methods in Health: Investigating Health and Health Services. New York, USA: McGraw-hill education; 2014.

75. Matthews B, Ross L. Research Methods: a Practical Guide for the Social Sciences. New York, USA: Pearson Education; 2010.
76. Shaughnessey J, Zechmeister E, Zechmeister J. Research Methods in Psychology. New York, USA: McGraw-Hill Education; 2014.
77. Shimada SL, Brandt CA, Feng H, McInnes DK, Rao SR, Rothendler JA, et al. Personal health record reach in the Veterans Health Administration: a cross-sectional analysis. J Med Internet Res 2014 Dec 12;16(12):e272 [FREE Full text] [doi: 10.2196/jmir.3751] [Medline: 25498515]
78. Andrews J, Shimp T. Advertising, Promotion, and Other Aspects of Integrated Marketing Communications. Tenth Edition. Boston, USA: Cengage Learning; 2017.
79. Kotler P, Armstrong G. Principles of Marketing. Upper Saddle River, USA: Pearson Higher Ed; 2017.
80. NHS login. 2019. URL: https://www.nhs.uk/using-the-nhs/nhs-services/nhs-login/ [accessed 2020-12-21]
81. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351. [doi: 10.2307/3250921]
82. Forquer HA, Christensen JL, Tan AS. Predicting continuance-findings from a longitudinal study of older adults using an eHealth newsletter. Health Commun 2014;29(9):937-946 [FREE Full text] [doi: 10.1080/10410236.2013.833580] [Medline: 24446900]
83. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011 Sep 30;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]

Abbreviations

ePHR: electronic personal health record
GP: general practice
NHS: National Health Service
SCR: summary care record

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by B McMillan, J Wang; comments to author 24.03.20; revised version received 22.04.20; accepted 14.05.20; published 13.01.21.

Please cite as:
Abd-Alrazaq A, Safi Z, Bewick BM, Househ M, Gardner PH
Patients' Perspectives About Factors Affecting Their Use of Electronic Personal Health Records in England: Qualitative Analysis
J Med Internet Res 2021;23(1):e17500
URL: http://www.jmir.org/2021/1/e17500/
doi:10.2196/17500
PMID:33439126

Original Paper

The Causal Relationship Between Portal Usage and Self-Efficacious Health Information–Seeking Behaviors: Secondary Analysis of the Health Information National Trends Survey Data

Jaeyoung Park[1], MSc; Muxuan Liang[2], DPhil; Jordan M Alpert[3], DPhil; Richard F Brown[4], DPhil; Xiang Zhong[1], DPhil

[1]Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, United States

[2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, United States

[3]Department of Advertising, University of Florida, Gainesville, FL, United States

[4]Department of Health Behavior and Policy, Virginia Commonwealth University School of Medicine, Richmond, VA, United States

Corresponding Author:
Xiang Zhong, DPhil
Department of Industrial and Systems Engineering
University of Florida
PO Box 115002
Gainesville, FL, 32611
United States
Phone: 1 3523922477
Email: xiang.zhong@ise.ufl.edu

Abstract

Background: Patient portals have drawn much attention, as they are considered an important tool for health providers in facilitating patient engagement. However, little is known about whether the intensive use of patient portals contributes to improved management of patients' health in terms of their confidence in acquiring health information and exercising self-care. There is a lack of randomized trials with these outcomes measured both pre- and postadoption of patient portals.

Objective: The aim of this study was to examine the causal relationship between the usage of patient portals and patients' self-efficacy toward obtaining health information and performing self-care.

Methods: This study was a secondary data analysis that used data from a US national survey, the National Cancer Institute's Health Information National Trends Survey 5 Cycle 1. Patient portal usage frequency was used to define the treatment. Survey items measuring self-efficacy on a Likert-type scale were selected as the main outcomes, including patients' confidence in obtaining health information and performing self-care. To establish causality using survey data, we adopted the instrumental variables method. To determine the direction of the causal relationship in the presence of high-dimensional confounders, we further proposed a novel testing framework that employs conditional independence tests in a directed acyclic graph. The average causal effect was measured using the two-stage least squares regression method.

Results: We showed that frequently using patient portals improves patients' confidence in obtaining health information. The estimand of the weighted average causal effect was 0.14 (95% CI 0.06-0.23; $P<.001$). This means that when increasing the portal usage intensity, for instance, from 1-2 times to 3-5 times per year, the expected average increase in confidence level measured on a Likert-type scale would be 0.14. However, we could not conclusively determine the causal effect between patient portal usage and patients' confidence in exercising self-care.

Conclusions: The results support the use of patient portals and encourage better support and education to patients. The proposed statistical method can be used to exploit the potential of national survey data for causal inference studies.

(J Med Internet Res 2021;23(1):e17782) doi:10.2196/17782

KEYWORDS

causal inference; instrumental variable; directed acyclic graph; patient portal; self-efficacy

Introduction

Given the growing evidence showing that patient engagement improves health outcomes and reduces health care costs, patient portals have drawn much attention. A patient portal is a secure online health platform linked to a patient's personal medical record that is 24/7 accessible from any location with an internet connection. Patient portals are considered an important tool for health providers in facilitating patient engagement [1,2]. Characteristics of portal users [3] and barriers to portal adoption [4,5] have been extensively investigated. Regarding the facilitators, it was found that patients who believed the patient portal was empowering demonstrated a higher intention to use [6]. The relationship between the internal use (quantified using engagement measurement) of online health platforms and external growth (its reach), and the social support activities related to users' participation has been examined as well [7,8]. Research has also focused on whether portal or secure messaging usage can affect the frequency of patients' office visits [9-15] and appointment adherence [16-19].

Nonetheless, little is known about whether patient portal usage contributes to the improved management of patients' health in terms of health literacy, communication, confidence in acquiring health information, and the self-monitoring and self-care of health. Interviews have indicated that patient's perception of access to online records are associated with a greater focus on their health and more proactive involvement in self-care [20,21]. However, there is a lack of randomized trials with self-efficacy outcomes measured both pre- and postportal adoption. Consequently, a causal link between patients' portal usage and patients' health information–seeking behaviors (self-efficacy) has not been formally established.

Despite the passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act and the recommendations by the US Institute of Medicine [22], the overall adoption rate of portals remains low [23]. A better understanding of the impact of patient portals and the benefit to patients is needed to increase adoption. To this end, we used the Health Information National Trends Survey (HINTS) data [24] to examine the causal relationship between the usage of patient portals and patients' self-efficacy in obtaining health information and performing self-care.

Methods

Study Setting

This study was a secondary data analysis that used data from the National Cancer Institute's HINTS 5 Cycle 1. HINTS 5 Cycle 1 is a cross-sectional survey of a nationally representative sample of US adults used to assess the impact of the health information environment. The survey was conducted from January 2017 through May 2017 using a self-administered mail questionnaire. Out of 10,265 surveys sent out, data were collected from 3285 (32% response rate) respondents [24]. The characteristics of these respondents are summarized in Table S1 in Multimedia Appendix 1.

Patients' self-efficacy in obtaining health information and performing self-care were considered as the outcomes. Survey items measuring patient self-efficacy were selected based on the Institute of Medicine's recommendations for promotion of patient portals to increase quality of care and reduce medical errors [22], as well as claims that health information technology, like portals, could increase patients' self-efficacy for managing their conditions [25-28]. Perceived confidence was assessed with the following questionnaire items: (1) "Overall, how confident are you that you could get advice or information about health or medical topics if you needed it?" (ConfidentGetHealthInfo, Y_1); (2) "Overall, how confident are you about your ability to take good care of your health? (OwnAbilityTakeCareHealth, Y_2)."

These measures were captured on a 5-point Likert-type scale where a higher score indicated greater confidence. Variables on patients' age, gender, race, ethnicity, marital status, education, employment status, household income, and insurance status were considered to be confounders in the study. Information on patients' portal activities was elicited by the following questionnaire item: "How many times did you access your online medical record in the last 12 months? (0/1 to 2 times/3 to 5 times/6 to 9 times/10 or more times)."

Different usage frequencies were considered to be different levels of treatment. For the instrumental variable (IV), we used the following questionnaire item: "Have any of your health care providers including doctors, nurses, or office staff ever encouraged you to use an online medical record? (yes/no)."

Missing values were sparse and were handled in several ways depending on the variable type: samples missing the outcomes were discarded; for the IVs, missing responses were replaced by "no encouragement"; and for the confounders, missing responses were imputed by the Multivariate Imputation by Chained Equations (MICE) [29] method with the IV and the confounders used as input variables for that method.

Establishing a Causal Relationship

We aimed to test the hypothesis that the exposure to patient portals, or intensively using a patient portal, will improve patients' self-efficacy outcomes. However, since we only used a one-time outcome measurement for each individual, we could not construct the before-after treatment contrast to directly measure the treatment effect. Further, without a randomized experimental design, the causality can be obscured by confounders. To address these issues, we adopted the IV method for causal inference in observational studies [30]. In these studies, IVs are used to adjust for both observed and unobserved confounding effects and to help identify a contrast of outcomes in the absence of temporal data.

We identified the following item as the binary variable and used it as the IV: "Have any of your health care providers including doctors, nurses, or office staff ever encouraged you to use an online medical record?"

Encouragement plays an influential role on patients' use of portals, as care providers' endorsement is an important factor in the adoption of these tools [6-8,31]. However, provider referrals to use patient portals have long been known to vary by race and socioeconomic status and are thought to be dependent on whether the provider believes that the patient will use the portal [32]. The same is true for the HINTS data we used, where income and education were observed to be associated with recommendation, patient portal usage, and self-efficacy. Thus, the requirement of no confounding between the IV and the outcome might not have been met if encouragement (or other candidate IVs, such as internet savviness) were used as an IV.

To further determine whether patient portal usage causes an improvement in self-efficacy, or vice versa, we proposed a testing framework that could both address the confounding issue and determine the direction of the causal relationship. Our testing framework was based on causal directed acyclic graphs (DAGs), which are used as a graphical tool to visually represent and understand the concepts of exposure, outcome, causation, and confounding [33]. To identify an appropriate IV, we generalized the criteria of IVs to allow for known (or observable) confounding among the IV, treatment, and outcome; and unknown (or unobservable) confounding between the treatment and outcome [30]. With this separation of confounders, only the known common confounders were essential for examining the causal relationship. To further identify the direction of causality, a testing framework employing conditional independence tests [34,35] was developed. The detailed description of the testing framework can be found in Multimedia Appendix 2.

Measuring Causal Effect

The proposed DAG-based testing framework aimed to qualitatively evaluate the causality. To quantify the treatment effect, two-stage least squares (TSLS) regression models were built. As the treatment (portal usage) has multiple levels (eg, 1-2 times and 3-5 times annually are different levels), the traditional average treatment effect is not identifiable. However, Angrist and Imbens [36] showed that an estimand, essentially a weighted average of per-unit average causal effects, is identifiable and can be estimated by TSLS regression models. The two major assumptions therein, independence and monotonicity of ordinal treatment effect, were thus naturally satisfied in our causal framework.

Results

Portal Enrollment

After removing the samples with too many missing data, we identified 3198 participants among the 3285 survey respondents. Among these 3198 participants, 1003 (31%) were self-reported patient portal users. For demographic and socioeconomic variables, the user group and the nonuser group had different characteristics (see Table S1 in Multimedia Appendix 1). Patients younger than 65 years old and females were more likely to be patient portal users. Moreover, participants who self-reported as White, married, and non-Hispanic were more likely to be users. In terms of income, higher income was positively correlated with a greater likelihood of being a portal user. Likewise, participants who were employed were more likely to use portals than those who were retired or unemployed. Compared to nonusers, portal users had a higher level of education (eg, undergraduate or postgraduate). Finally, compared to those covered by employer-provided insurance, patients who were covered by private insurance, Medicare, or Medicaid were less likely to enroll in patient portals.

Encouragement and Portal Usage

Next, we characterized the users' portal usage behavior. Of the 1003 portal users, 49% (496) reported using portals 1-2 times in the past 12 months, 31% (313) reported using them 3-5 times, 10% (104) reported using them 6-9 times, and 9% (90) reported using them more than 9 times.

Of the 3198 participants, 1375 (43%) were encouraged to use patient portals and 1823 (57%) were not. Among the 1375 respondents who were encouraged, 549 (40%), 383 (28%), 267 (19%), and 176 (13%) individuals never used a portal, used a portal 1-2 times, 3-5 times, and more than 5 times, respectively. In contrast, there were 1646 (90%), 113 (6%), 46 (3%), and 18 (1%) participants in the nonencouraged group, respectively. It was evident that the IV and the treatment were significantly associated, which was verified by a chi-square test ($P<.001$).

Encouragement and Self-Efficacy

There were more patients with positive responses for the self-efficacy outcomes in the patient population who were recommended to use portals. The distributions of each outcome variable conditioning on the value of the IV are displayed in Table 1, and chi-square test results are provided. Among the 3198 participants, 3111 and 3165 participants were identified as those who answered questions Y1 and Y2, respectively. For ConfidentGetHealthInfo, individuals in the encouraged group (G_e) were more likely to be completely confident or very confident than those in the nonencouraged group (G_n; G_e: 907/1353, 67%; G_n: 1000/1758, 57%; $P<.001$). For OwnAbilityTakeCareHealth, G_e patients were slightly more confident in self-care (G_e: 998/1364, 73%; G_n: 1222/1801, 68%; $P=.001$). All the comparisons suggest that the IV and the outcomes were significantly associated.

Table 1. Chi-square test results for the association between encouragement to use patient portals and self-efficacy outcomes.

Outcomes	IV[a]: encouragement to use patient portals		*P* value
	Yes, n (%)	No, n (%)	
ConfidentGetHealthInfo, Y_1	1353 (100)	1758 (100)	<.001
Not confident at all	21 (1.5)	50 (2.8)	
A little confident	56 (4.1)	108 (6.1)	
Somewhat confident	369 (27.3)	600 (34.1)	
Very confident	566 (41.8)	607 (34.5)	
Completely confident	341 (25.2)	393 (22.4)	
OwnAbilityTakeCareHealth, Y_2	1364 (100)	1801 (100)	.001
Not confident at all	15 (1.1)	31 (1.7)	
A little confident	31 (2.3)	79 (4.4)	
Somewhat confident	320 (23.5)	469 (26.0)	
Very confident	652 (47.8)	815 (45.3)	
Completely confident	346 (25.4)	407 (22.6)	

[a] IV: instrumental variable.

Causal Relationship Between Patient Portal Usage and Self-Efficacy

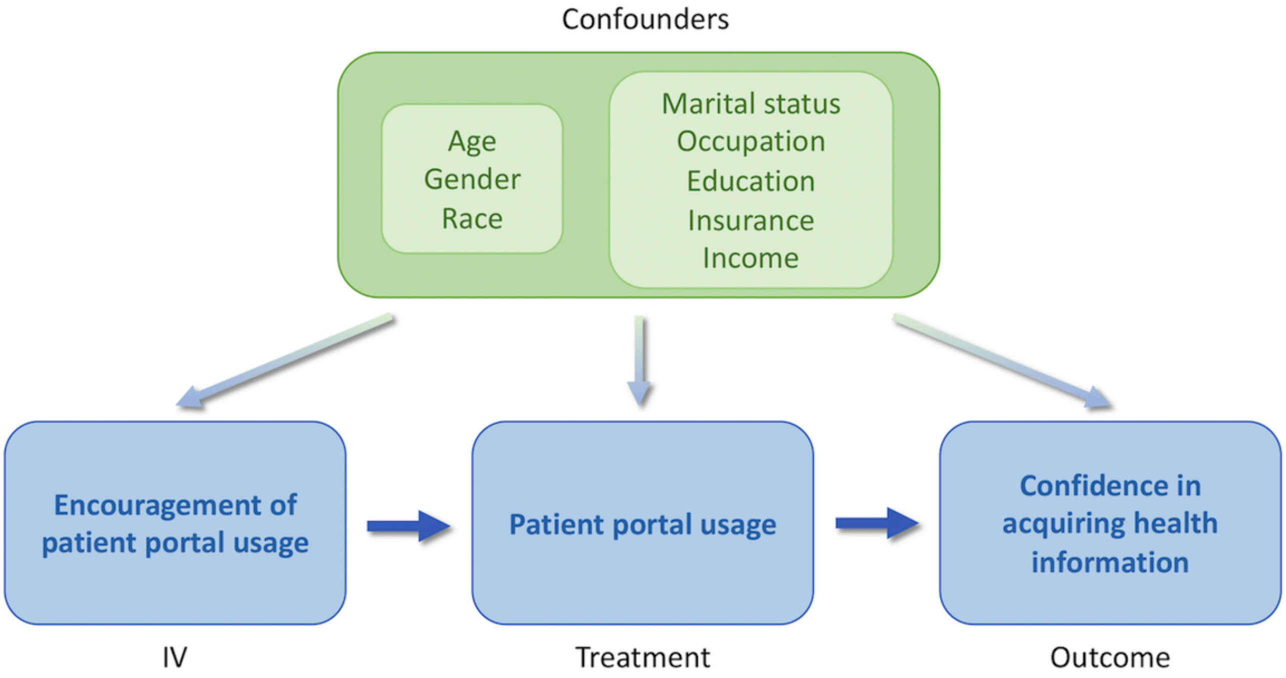
Following the testing procedure described in Multimedia Appendix 2, conditional independence tests were conducted. Test A tested the hypothesis that, given the common confounders and the treatment, the IV is conditionally independent with the outcome. This test examined whether higher self-efficacy is a cause for increased portal usage. Test B tested the hypothesis that the IV and the outcome are conditionally independent, given the common confounders alone. This test examined whether the increase in portal usage is a cause for higher self-efficacy. A *P* value <.05 was used as the statistical significance level for the overall test. The results are shown in Table 2.

Table 2. Results of conditional independence tests.

Outcomes	*P* value	
	Test A: $Z \perp Y \mid$	Test B: $Z \perp Y \mid$
ConfidentGetHealthInfo	.28	.02
OwnAbilityTakeCareHealth	.42	.23

For ConfidentGetHealthInfo, we could not reject the hypothesis for Test A (overall *P*=.28) but could reject Test B (overall *P*=.02), meaning that portal usage did have a causal effect on self-efficacy toward obtaining health information. This relationship is illustrated in Figure 1. However, for OwnAbilityTakeCareHealth, we could not reject the hypothesis for either Test A (overall *P*=.42) or Test B (overall *P*=.23). In this case, we could not determine the causal relationship between patients' portal usage and patients' confidence in exercising self-care.

Figure 1. The causal relationship among encouragement, portal usage, and confidence in seeking health information with the confounders. IV: instrumental variable.



Treatment Effect of Patient Portal Usage on Self-Efficacy

A TSLS regression model was built to quantify the treatment effect of portal usage on self-efficacy toward acquiring health information. The estimand of the weighted average causal effect was 0.14 (95% CI 0.06-0.23; *P*<.001). This indicates that, when increasing the portal usage intensity, for instance, from 1-2 times to 3-5 times per year, the expected average increase in the reported confidence level (measured with a Likert-type scale) would be 0.14. It can be concluded that the more intensely patients engage with portals, the more confident they are in obtaining health information.

Discussion

Identifying Causality

Existing work on survey analysis mainly focuses on the strength of the association between questionnaire items and targeted outcomes [37,38], as there is a lack of understanding regarding the causal path diagram and a lack of methods for determining if the path diagram is congruent with the data. In this study, we developed a new framework that allowed us to detect causality and estimate the magnitude of the treatment effect represented in the path diagram using the HINTS 5 survey. Our framework was able to eliminate the estimation bias due to unmeasured confounders, and guarantee the test efficiency given the limited sample size of the survey. Moreover, the testing framework was robust to the choice of IVs. In addition to using "encouragement" as an IV, we tested internet savviness as an alternative IV and found the conclusion was consistent (see Multimedia Appendix 2).

Effect of Portal Usage

Although we should be cautious in interpreting the relationship between using patient portals and self-efficacy as causal, we did observe a co-occurrence of better outcomes and increased patient portal usage intensity. Despite the benefits of patient portals being well documented [39,40], rates of usage remain low [41], and increasing the adoption of patient portals would provide more patients a trusted source of personalized information. Because patients can be easily misinformed by the large amount of inaccurate information online [42], using patient portals allows patients and their caregivers to stay connected with their providers. This occurs through easy access to their own health information and the ability to contact providers via secure messaging if they have questions or concerns. Our framework further confirmed that portal usage does positively affect confidence in obtaining health information. However, we did not observe a similar level of significance to conclude that higher portal usage will lead to higher self-efficacy in performing self-care. It is possible that other confounding factors, such as income and education levels [43], might have had a more powerful effect on the outcome.

Encouraging Patients' Portal Use

Our analysis shows that being encouraged to use patient portals positively affects the intensity of portal usage, which in turn influences people's confidence in acquiring health information. Other studies have reported that actively using patient portals, such as sending messages to physicians and viewing prescription and lab results, is positively associated with high-quality physician-patient relationships and patients' confidence to understand health information. It is evident that a trusting physician-patient relationship helps promote healing and remove barriers to obtaining medical information [44]. However, efforts to increase the adoption of patient portals often rely on pamphlets and flyers [45] or hurried conversations by medical staff. Given the potential benefits of patient portals and our findings that encouragement affects usage, we believe that an innovative intervention is necessary to increase patient adoption and usage, which ultimately can lead to significant improvements in patient outcomes.

Limitations

The study was conducted using HINTS 5 Cycle 1 data, which were collected in 2017. The study can be improved by combining multiple data sets, including ones published recently [46]. The power of the test can be increased by including more samples, which can potentially help identify the causal relationship between portal usage and self-care. It is worth noting that survey designs are not identical across all three cycles of HINTS 5. For instance, the outcome of ConfidenceGetHealthInfo is available in Cycle 1 (2017) and Cycle 3 (2019) but not in Cycle 2 (2018). Considerable effort will be needed to merge these data sets. Furthermore, the HINTS database contains rich information beyond the variables that were used for this study, and other aspects (eg, health disparities) related to both portal usage and self-efficacy behaviors should be explored.

The establishment of the causal relationship between portal usage and patients' self-efficacy demands that encouragement to use portals is not based on patients' self-efficacy. We have observed that many health care organizations have integrated the facilitation of portal enrollment into their new patients' registration protocol [47]. In addition, portal use encouragement also occurs during a patient's interaction with the front desk when addressing billing and appointment scheduling, as patient portals can be an alternative venue to handle these services. These scenarios correspond to the cases that nurses or office staff encourage patients to use an online medical record. These situations account for the majority of the encouragement during patient encounters and thus support the selection of encouragement as an IV. However, as for personal encouragement from a physician, we are unaware of any health care organizations that have policies or incentives for providers to systematically encourage all patients to use portals, meaning that encouragement can be at the providers' discretion [48]. Therefore, collecting physicians' input concerning their patterns of encouragement would be valuable for further justifying the validity of using encouragement as an IV. To ensure the rigor of our results, we further tested internet savviness as an alternative IV and found a consistent conclusion (see Multimedia Appendix 2).

Although the confounders between the treatment and the outcome can be completely unmeasured, we still require the common confounders to be fully observable. The identification of common confounders largely relies on domain knowledge. It is worth noting that the choice does not have to be unique. In our casual diagram, adding variables to the common confounder set still results in a valid choice. However, chi-square tests can be sensitive to the choice of confounders when the sample size is not sufficiently large [49]. Therefore, to make the results more robust, more survey respondents are needed.

Conclusions

Since the Affordable Care Act mandated portal usage, enthusiasm for portals has declined; however, this study found that using patient portals improves patients' confidence in obtaining advice or information about health or medical topics. Our findings thus attest to the benefit of patient portals and to providing better support and education to patients. In addition, our proposed statistical method exploits the potential of using national survey data such as the HINTS program to examine causal effects to obtain new insights. Theoretically, the treatment effect can be heterogeneous based on different patient characteristics. There is thus a need to develop a testing framework that can identify the disparity in causal effects. For justifying the clinical insights identified in this study, we cannot solely rely on patients' self-reported outcomes, but should also survey physicians on their perception of patients' self-efficacy. Furthermore, how physicians make encouragement decisions should also be investigated, and a randomized controlled study with pre- and posttreatment outcomes being clearly documented is necessary to fully understand the treatment effect on self-efficacy outcomes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1.
[PDF File (Adobe PDF File), 56 KB - jmir_v23i1e17782_app1.pdf]

Multimedia Appendix 2
Methods.
[DOCX File , 147 KB - jmir_v23i1e17782_app2.docx]

References

1. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

2. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary care providers' views of patient portals: interview study of perceived benefits and consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
3. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
4. Kooij L, Groen WG, van Harten WH. Barriers and facilitators affecting patient portal implementation from an organizational perspective: qualitative study. J Med Internet Res 2018 May 11;20(5):e183 [FREE Full text] [doi: 10.2196/jmir.8989] [Medline: 29752253]
5. Scott KC, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]
6. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
7. Gopalsamy R, Semenov A, Pasiliao E, McIntosh S, Nikolaev A. Engagement as a driver of growth of online health forums: observational study. J Med Internet Res 2017 Aug 29;19(8):e304 [FREE Full text] [doi: 10.2196/jmir.7249] [Medline: 28851677]
8. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. J Med Internet Res 2017 Apr 24;19(4):e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]
9. Palen TE, Ross C, Powers JD, Xu S. Association of online patient access to clinicians and medical records with use of clinical services. JAMA 2012 Nov 21;308(19):2012-2019. [doi: 10.1001/jama.2012.14126] [Medline: 23168824]
10. Otte-Trojel T, de BA, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul;21(4):751-757. [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
11. Zhou YY, Leith WM, Li H, Tom JO. Personal health record use for children and health care utilization: propensity score-matched cohort analysis. J Am Med Inform Assoc 2015 Jul;22(4):748-754. [doi: 10.1093/jamia/ocu018] [Medline: 25656517]
12. Leveille SG, Mejilla R, Ngo L, Fossa A, Elmore JG, Darer J, et al. Do patients who access clinical information on patient internet portals have more primary care visits? Med Care 2016 Jan;54(1):17-23. [doi: 10.1097/MLR.0000000000000442] [Medline: 26565525]
13. Bavafa H, Hitt LM, Terwiesch C. The impact of e-visits on visit frequencies and patient health: evidence from primary care. Manage Sci 2018 Dec;64(12):5461-5480 [FREE Full text] [doi: 10.1287/mnsc.2017.2900] [Medline: 33033417]
14. Bergmo TS, Kummervold PE, Gammon D, Dahl LB. Electronic patient-provider communication: will it offset office visits and telephone consultations in primary care? Int J Med Inform 2005 Sep;74(9):705-710. [doi: 10.1016/j.ijmedinf.2005.06.002] [Medline: 16095961]
15. Meng D, Palen TE, Tsai J, McLeod M, Garrido T, Qian H. Association between secure patient-clinician email and clinical services utilisation in a US integrated health system: a retrospective cohort study. BMJ Open 2015 Nov 09;5(11):e009557 [FREE Full text] [doi: 10.1136/bmjopen-2015-009557] [Medline: 26553841]
16. Horvath M, Levy J, L'Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. J Med Internet Res 2011;13(2):e41 [FREE Full text] [doi: 10.2196/jmir.1702] [Medline: 21616784]
17. Lin C, Mistry N, Boneh J, Li H, Lazebnik R. Text message reminders increase appointment adherence in a pediatric clinic: a randomized controlled trial. Int J Pediatr 2016;2016:8487378 [FREE Full text] [doi: 10.1155/2016/8487378] [Medline: 28127311]
18. Mendel A, Chow S. Impact of health portal enrolment with email reminders at an academic rheumatology clinic. BMJ Qual Improv Rep 2017;6(1):1-9 [FREE Full text] [doi: 10.1136/bmjquality.u214811.w5926] [Medline: 28321302]
19. Zhong X, Park J, Liang M, Shi F, Budd PR, Sprague JL, et al. Characteristics of patients using different patient portal functions and the impact on primary care service utilization and appointment adherence: retrospective observational study. J Med Internet Res 2020 Feb 25;22(2):e14410 [FREE Full text] [doi: 10.2196/14410] [Medline: 32130124]
20. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
21. Dhanireddy S, Walker J, Reisch L, Oster N, Delbanco T, Elmore JG. The urban underserved: attitudes towards gaining full access to electronic medical records. Health Expect 2014 Oct;17(5):724-732. [doi: 10.1111/j.1369-7625.2012.00799.x] [Medline: 22738155]
22. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington (DC): National Academies Press (US); 2001:A. [doi: 10.17226/10027] [Medline: 25057539]
23. Fraccaro P, Vigo M, Balatsoukas P, Buchan IE, Peek N, van der Veer SN. Patient portal adoption rates: a systematic literature review and meta-analysis. Stud Health Technol Inform 2017;245:79-83. [Medline: 29295056]

24. About HINTS. The Health Information National Trends Survey. URL: https://hints.cancer.gov/about-hints/learn-more-about-hints.aspx [accessed 2019-08-29]
25. Dorr D, Bonner LM, Cohen AN, Shoai RS, Perrin R, Chaney E, et al. Informatics systems to promote improved care for chronic illness: a literature review. J Am Med Inform Assoc 2007;14(2):156-163 [FREE Full text] [doi: 10.1197/jamia.M2255] [Medline: 17213491]
26. Rigby M, Georgiou A, Hyppönen H, Ammenwerth E, de Keizer N, Magrabi F, et al. Patient portals as a means of information and communication technology support to patient-centric care coordination - the missing evidence and the challenges of evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL. Yearb Med Inform 2015 Aug 13;10(1):148-159 [FREE Full text] [doi: 10.15265/IY-2015-007] [Medline: 26123909]
27. Allen A, Des Jardins TR, Heider A, Kanger CR, Lobach DF, McWilliams L, et al. Making it local: beacon communities use health information technology to optimize care management. Popul Health Manag 2014 Jun;17(3):149-158 [FREE Full text] [doi: 10.1089/pop.2013.0084] [Medline: 24476558]
28. McCarrier KP, Ralston JD, Hirsch IB, Lewis G, Martin DP, Zimmerman FJ, et al. Web-based collaborative care for type 1 diabetes: a pilot randomized trial. Diabetes Technol Ther 2009 Apr;11(4):211-217 [FREE Full text] [doi: 10.1089/dia.2008.0063] [Medline: 19344195]
29. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2010;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]
30. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med 2014 Jun 15;33(13):2297-2340 [FREE Full text] [doi: 10.1002/sim.6128] [Medline: 24599889]
31. Shah SD, Liebovitz D. It takes two to tango: engaging patients and providers with portals. PM&R 2017 May;9(5S):S85-S97. [doi: 10.1016/j.pmrj.2017.02.005] [Medline: 28527507]
32. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
33. Pearl J. Causal diagrams for empirical research. Biometrika 1995;82(4):669-688. [doi: 10.1093/biomet/82.4.669]
34. Conditional Independence. PennState. URL: https://online.stat.psu.edu/stat504/lesson/5/5.4/5.4.4
35. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
36. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American Statistical Association 1995 Jun;90(430):431-442. [doi: 10.1080/01621459.1995.10476535]
37. Hou J, Shim M. The role of provider-patient communication and trust in online sources in internet use for health-related activities. J Health Commun 2010;15 Suppl 3:186-199. [doi: 10.1080/10810730.2010.522691] [Medline: 21154093]
38. Hong H, Oh HJ. The effects of patient-centered communication: exploring the mediating role of trust in healthcare providers. Health Commun 2020 Apr;35(4):502-511. [doi: 10.1080/10410236.2019.1570427] [Medline: 30706741]
39. Reed ME, Huang J, Brand RJ, Neugebauer R, Graetz I, Hsu J, et al. Patients with complex chronic conditions: health care use and clinical events associated with access to a patient portal. PLoS One 2019;14(6):e0217636 [FREE Full text] [doi: 10.1371/journal.pone.0217636] [Medline: 31216295]
40. Stewart MT, Hogan TP, Nicklas J, Robinson SA, Purington CM, Miller CJ, et al. The promise of patient portals for individuals living with chronic illness: qualitative study identifying pathways of patient engagement. J Med Internet Res 2020 Jul 17;22(7):e17744 [FREE Full text] [doi: 10.2196/17744] [Medline: 32706679]
41. Tsai R, Bell EJ, Woo H, Baldwin K, Pfeffer MA. How patients use a patient portal: an institutional case study of demographics and usage patterns. Appl Clin Inform 2019 Jan;10(1):96-102 [FREE Full text] [doi: 10.1055/s-0038-1677528] [Medline: 30727003]
42. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41:433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
43. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 2003 May;14(3):300-306. [Medline: 12859030]
44. Thom DH, Hall MA, Pawlson LG. Measuring patients' trust in physicians when assessing quality of care. Health Aff (Millwood) 2004;23(4):124-132 [FREE Full text] [Medline: 15318572]
45. Bronsburg S, Durity M, Ramim M. An exploratory study of patient portal adoption and use in multiple clinical specialty departments. OJAKM 2019 Oct 10;7(1):86-101. [doi: 10.36965/ojakm.2019.7(1)86-101]
46. HINTS 5 overview and schedule. The Health Information National Trends Survey. URL: https://hints.cancer.gov/about-hints/data-collection-schedule.aspx [accessed 2020-08-29]
47. Ancker JS, Nosal S, Hauser D, Way C, Calman N. Access policy and the digital divide in patient access to medical records. Health Policy and Technology 2017 Mar;6(1):3-11. [doi: 10.1016/j.hlpt.2016.11.004]
48. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870. [doi: 10.1093/jamia/ocz023] [Medline: 30958532]

49. McHugh ML. The chi-square test of independence. Biochem Med (Zagreb) 2013;23(2):143-149. [Medline: 23894860]

Abbreviations

DAG: directed acyclic graph
HINTS: Health Information National Trends Survey
HITECH: Health Information Technology for Economic and Clinical Health
IV: instrumental variable
MICE: Multivariate Imputation by Chained Equations
TSLS: two-stage least squares

Edited by G Eysenbach, Q Zeng; submitted 14.01.20; peer-reviewed by A Nikolaev, N Benda, W van Harten, D Pförringer, V Osadchiy, C Luckhurst; comments to author 28.08.20; revised version received 25.09.20; accepted 21.12.20; published 27.01.21.

Please cite as:
Park J, Liang M, Alpert JM, Brown RF, Zhong X
The Causal Relationship Between Portal Usage and Self-Efficacious Health Information–Seeking Behaviors: Secondary Analysis of the Health Information National Trends Survey Data
J Med Internet Res 2021;23(1):e17782
URL: http://www.jmir.org/2021/1/e17782/
doi:10.2196/17782
PMID:33502334

Original Paper

Application of an Adaptive, Digital, Game-Based Approach for Cognitive Assessment in Multiple Sclerosis: Observational Study

Wan-Yu Hsu[1], PhD; William Rowles[1], BA; Joaquin A Anguera[1,2,3], PhD; Chao Zhao[1], MSc; Annika Anderson[1], BA; Amber Alexander[1], BSc; Simone Sacco[1], MD; Roland Henry[1], PhD; Adam Gazzaley[1,2,3,4], MD, PhD; Riley Bove[1], MD

[1]Department of Neurology, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States
[2]Neuroscape, University of California, San Francisco, San Francisco, CA, United States
[3]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, United States
[4]Department of Physiology, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Riley Bove, MD
Department of Neurology
Weill Institute for Neurosciences
University of California, San Francisco
675 Nelson Rising Lane
San Francisco, CA, 94158
United States
Phone: 1 415 595 2795
Email: riley.bove@ucsf.edu

Related Article:

This is a corrected version. See correction statement: http://jmir.org/2021/1/e27440/

Abstract

Background: Cognitive impairment is one of the most debilitating manifestations of multiple sclerosis. Currently, the assessment of cognition relies on a time-consuming and extensive neuropsychological examination, which is only available in some centers.

Objective: To enable simpler, more accessible cognitive screening, we sought to determine the feasibility and potential assessment sensitivity of an unsupervised, adaptive, video game–based digital therapeutic to assess cognition in multiple sclerosis.

Methods: A total of 100 people with multiple sclerosis (33 with cognitive impairment and 67 without cognitive impairment) and 24 adults without multiple sclerosis were tested with the tablet game (EVO Monitor) and standard measures, including the Brief International Cognitive Assessment for Multiple Sclerosis (which included the Symbol Digit Modalities Test [SDMT]) and Multiple Sclerosis Functional Composite 4 (which included the Timed 25-Foot Walk test). Patients with multiple sclerosis also underwent neurological evaluations and contributed recent structural magnetic resonance imaging scans. Group differences in EVO Monitor performance and the association between EVO Monitor performance and standard measures were investigated.

Results: Participants with multiple sclerosis and cognitive impairment showed worse performance in EVO Monitor compared with participants without multiple sclerosis (P=.01) and participants with multiple sclerosis without cognitive impairment (all P<.002). Regression analyses indicated that participants with a lower SDMT score showed lower performance in EVO Monitor (r=0.52, P<.001). Further exploratory analyses revealed associations between performance in EVO Monitor and walking speed (r=–0.45, P<.001) as well as brain volumetric data (left thalamic volume: r=0.47, P<.001; right thalamic volume: r=0.39, P=.002; left rostral middle frontal volume: r=0.28, P=.03; right rostral middle frontal volume: r=0.27, P=.03).

Conclusions: These findings suggest that EVO Monitor, an unsupervised, video game–based digital program integrated with adaptive mechanics, is a clinically valuable approach to measuring cognitive performance in patients with multiple sclerosis.

Trial Registration: ClinicalTrials.gov NCT03569618; https://clinicaltrials.gov/ct2/show/NCT03569618

(J Med Internet Res 2021;23(1):e24356) doi:10.2196/24356

KEYWORDS

cognition; digital health; mHealth; multiple sclerosis; cognitive assessment; video game

Introduction

Cognitive impairment (CI) occurs in 30% to 70% of people with multiple sclerosis (MS) and has a profound influence on a patient's personal functioning, social interaction, employment, and overall quality of life [1,2]. Being able to effectively detect CI is essential to better managing further decline [3] and helping patients navigate problems related to their daily living. The most commonly affected cognitive domains are processing speed, attention, executive function, and memory [4,5]. Currently, clinical cognitive assessment relies on a comprehensive neuropsychological examination, which is time-consuming and extensive. The examination results may be affected by patient fatigue or loss of engagement. Given the large interindividual variability in the pattern of CI in MS [6,7], the traditional nonadaptive assessments may overlook cognitive deficiencies [8]. Taking advantage of technology, digital tools offer a platform to integrate personalizing features, including adaptive staircase algorithms to video game–style mechanics for cognitive assessment [8-15]. This approach can mitigate potential ceiling and floor effects when interindividual variability is high, and leads to more reliable assessments that can be completed in a timely manner [8,16]. Furthermore, digital tools can be easily applied in different settings, including patients' homes, which substantially improves health care accessibility for patients who have difficulties with travel to the clinic due to cognitive or physical disabilities [14,17,18].

Digital tools have been used for functional assessment, rehabilitation, and health care monitoring in clinical populations such as stroke [19,20], schizophrenia [21], depression [10,12], attention-deficit/hyperactivity disorder [15,22-24], and neurodevelopmental disorders [8,9]. In MS, studies have demonstrated that by using digital tools, health care services can be delivered effectively [25,26] and even that comprehensive neurological exams can be performed remotely [27]. However, to the best of our knowledge, applications of serious video game–based digital tools incorporating closed-loop adaptation mechanics as cognitive assessments for people with MS have not been assessed.

We previously evaluated whether a video game–based digital tool, EVO-AKL-T01 (Akili), could improve cognition in MS [17,18]. After these studies, we looked to better characterize cognitive function using similar tools based on these cognitive therapeutics. EVO Monitor (Akili) was developed based on findings that a precursor (NeuroRacer), embedded with adaptive algorithms, is sensitive to age-related cognitive decline across the lifespan and can enhance cognitive control [28]. EVO Monitor, which includes some of NeuroRacer's early closed-loop features, is a novel, tablet-based, digital platform that is incorporated with video game mechanisms, visual and auditory feedback, adaptive algorithms, and sophisticated graphics that is designed to assess executive function, attention, and information processing speed for clinical populations. Specifically, it consists of 3 tasks: perceptual discrimination, visuomotor tracking, and multitasking. In the perceptual discrimination task, the participants complete a go/no go–like paradigm, in which they tap the iPad screen for correctly colored target stimuli while ignoring distracting targets. The visuomotor tracking task requires the participant to tilt the iPad to steer an avatar around obstacles. The multitasking task requires participants to perform both perceptual discrimination and visuomotor tracking concurrently. The tool is designed to challenge attention, goal management, and information processing speed in the setting of interference. EVO Monitor has been validated as a tool to assess cognitive ability in children with and without neurodevelopmental disabilities. For example, it can differentiate between the performance of children with or without 16p11.2 BP4-BP5 deletion and healthy controls, whereas traditional nonadaptive cognitive assessments overlook group differences [8].

In this study, we aimed to determine the feasibility and potential assessment sensitivity of EVO Monitor as an unsupervised, tablet game–based approach with adaptive algorithms to assess cognition in MS. In this context, 100 people with MS and 24 adults without MS were tested with EVO Monitor and standard measures, including the written version of Symbol Digit Modalities Test (SDMT), the most sensitive test for detecting cognitive involvement in the MS course [29,30]. Group-level differences were examined to evaluate whether EVO Monitor is sensitive to differences among participants with MS with and without CI as well as participants without MS. The association between performance in EVO Monitor and standard measures was investigated. Furthermore, since cognition in MS has been shown to correlate with physical fitness, where aerobic capacity and muscular strength outcomes are associated with cognitive processing speed and inhibitory control [31,32], and magnetic resonance imaging (MRI) measures have shown that thalamus damage is associated with the presence of CI, and frontal lesion is associated with executive function [33,34], exploratory analyses of the correlation between performance in EVO Monitor and both physical measures and MRI volumetric data (thalamus and frontal lobes) were conducted to understand whether cognitive performance assessed by a video game–based digital tool (EVO Monitor) would be associated with physical activity and structural MRI measures.

Methods

Participants

A total of 100 adults with a diagnosis of clinically isolated syndrome or MS [35] (mean age: 52.2, standard error of the mean [SEM] 1.24 years) were recruited from the University of California, San Francisco (UCSF) Multiple Sclerosis and Neuroinflammation Center. Patients with clinical relapses or steroid use in the past month or with severe visual, cognitive, or motor impairment that would preclude the use of a tablet-based tool were excluded. A group of 24 adults without MS (non-MS) (mean age: 46.0, SEM 3.72 years) with no chronic autoimmune diseases were also recruited from the UCSF staff, willing family members of patients in the clinic, and other eligible and willing volunteers. All participants with MS were recruited as part of studies to determine the feasibility [18] and preliminary efficacy [17] of the EVO platform as a digital therapeutic to improve processing speed in people with MS. The analysis of this study was based on baseline performance data (ie, before any cognitive intervention) of our feasibility [18] and efficacy [17] trials. All procedures performed in the study involving human participants were approved by the Committee for Human Research at the University of California, San Francisco. Written informed consent was obtained from each participant. The trial is registered with ClinicalTrials.gov (NCT03569618).

Task Description

Standard Measures

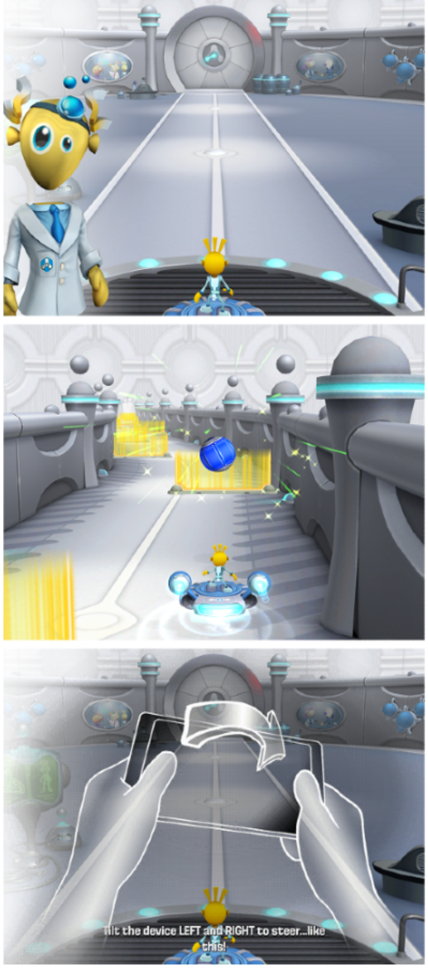
The Brief International Cognitive Assessment for Multiple Sclerosis is a cognitive assessment tool that is validated in MS populations as compared with participants without MS [36]. It is a standardized, internationally validated battery [37] including (1) the SDMT, a widely used measure of attention and information processing speed in MS [29,30]; (2) the California Verbal Learning Test Second Edition, a verbal memory immediate recall test [38]; and (3) the Brief Visuospatial Memory Test Revised, a visual memory immediate recall test [39]. Serial versions of all tests were used to minimize practice effects [40].

The MS Functional Composite 4 evaluates 4 key MS-related functional domains [41]: walking speed (Timed 25-Foot Walk [T25FW]), upper extremity function (Nine-Hole Peg Test), vision (Sloan low-contrast letter acuity test), and cognition (with a test of information processing speed, attention, and working memory [Paced Auditory Serial Addition Task] [42]).

EVO Monitor

EVO Monitor is a digital cognitive assessment developed by Akili Interactive Labs (Akili, Boston) to assess cognitive function, including attention and related cognitive control processes in clinical populations (Figure 1). The program is an immersive action video game that has been engineered with adaptive algorithms to target fronto-parietal brain networks fundamentally linked to attentional control among other aspects of cognition. It was developed based on the principles of NeuroRacer, an innovative cognitive intervention that is sensitive to age-related cognitive decline [28]. EVO Monitor comprises 3 tasks: perceptual discrimination, visuomotor tracking, and multitasking. In the perceptual discrimination task, the participants are instructed to respond to colored target stimuli by tapping the iPad screen while ignoring distractors. In the visuomotor tracking task, the participants navigate a character along a dynamically moving road while avoiding walls and obstacles by tilting the iPad. The multitasking task requires participants to perform both perceptual discrimination and visuomotor tracking at the same time until participants complete a minimum number of trials and reach a stable level of performance.

Figure 1. Screenshot of the EVO Monitor cognitive assessment program. The participants are instructed to respond to colored target stimuli by tapping the iPad screen while navigating a character along a dynamically moving road and avoiding walls and obstacles by tilting the iPad. Copyright © 2020-2021, Akili Interactive Labs, Inc. All rights reserved.



The adaptive algorithms change game difficulty on a trial-by-trial basis for both the perceptual discrimination (adapting the response window for a target) and visuomotor tracking (adapting the speed of the forward path), with real-time feedback making the participants aware of their performance. More specifically, the closed-loop adaptive algorithm makes proportional changes in gameplay difficulty to keep the player's performance at an approximate 80% rate of accuracy based on adaptive psychometric principles [43-45], which ensures that task difficulty is equated across participants [45,46] and enhances participant engagement [47]. It takes approximately 7 minutes to complete the tasks, with a specified number of both correct and incorrect trials allowing the adaptive algorithm to settle on a prescribed level of difficulty (ie, threshold level) that would converge on a consistent accuracy rate. The threshold level represents task performance, as it indicates the task difficulty at which the participants achieve approximately an 80% rate of accuracy.

At the beginning of the EVO Monitor session, a brief instruction is given on the iPad screen with a short practice to ensure the participant understands the tasks, followed by one session of the actual assessment, which takes about 7 minutes, with a task order of multitasking, perceptual discrimination, visuomotor tracking, and multitasking. Each task is about 1.5 to 2 minutes with no break in between. Although the assessment is self-guided, a study coordinator sat in with participants to ensure they were following the instructions. None of the participants had played EVO Monitor before participation. Since EVO Monitor continuously monitors the user's performance at a rate of 30 frames per second, the measured reaction time, perceptual discrimination task sensitivity (eg, hit and false alarm rates), and visuomotor tracking performance generate 39 basic performance metrics. Among the 39 metrics extracted from participants' navigation of EVO Monitor, we prespecified the calculated threshold levels during both single (perceptual discrimination, visuomotor tracking) and multitasking conditions as attentional measures, according to previous studies [8,16].

Basic Reaction Time

To ensure that any observed differences in cognitive measures between groups were not due to differences in motoric speed, we assessed basic response speed of participants on a simple task with minimal loading on executive function skills [28]. This task was designed to index the motoric speed, and the measured data were only included in the analyses as a covariate to control for potential motor speed deficits in participants with MS. In this task, participants were instructed to respond to a target stimulus (40 trials) as fast as they could by tapping a button on an iPad platform. Similar to EVO Monitor, this task used adaptive algorithms that modulate the challenge level of the task on a trial-by-trial basis based on individual performance.

Only data from the dominant hand were included as each participant's basic reaction time (BRT) in the following analyses.

Clinically Acquired MRI Measures

Clinically or research-acquired brain isotropic T1 and T2 fluid-attenuated inversion recovery images were available for 56 participants with MS (16/56, 29% with 1.5-T and 40/56, 71% with 3-T MRIs) at a mean of 76.0 (SEM 33.9) days before the study visit. Lesion segmentation was performed using the LST (lesion segmentation toolbox) lesion probability algorithm 2.0 DICOM (Digital Imaging and Communications in Medicine) v1.4 segmentation pipeline, which creates lesion probability maps, masks, and labels. These were then manually validated by an expert radiologist (SS). Volumetric analysis was performed from T1 anatomical images using 3 complementary tools: FreeSurfer 5.3 and ANTs Morphology 2.1.0 [48], used to segment tissue into cerebrospinal fluid, cortical grey matter, subcortical grey matter, white matter, brainstem, and cerebellum, and Mindboggle 1.0 [49], which combines the morphology outputs of FreeSurfer and ANTs to generate volume images and tabular information for further analysis. Bilateral thalamic and frontal lobe volumetric measures were normalized to individual intracranial volume [50] for exploratory analyses of association between performance in EVO Monitor and MRI volumetric measures. The selection of the thalamus and frontal lobe as regions of interest was based on ample evidence indicating an association between atrophy in these regions and cognitive dysfunction in MS [33,34].

Statistical Analysis

In order to evaluate whether EVO Monitor is sensitive to differences among participants with MS with and without CI as well as participants without MS, participants with MS were divided into 2 subgroups (ie, CI and non-CI) according to their baseline SDMT *z* scores. We characterized the participants with a SDMT *z* score lower than –1 based on published normative data [51] as having CI. Group-level differences were assessed with one-way analysis of covariance (ANCOVA) with age, sex, years of education, and BRT as covariates to control for potential differences in demographic features and motoric quickness. Two-tailed Student *t* tests were carried out for post hoc comparisons when appropriate. To discern the association between EVO Monitor and standard measures, Pearson correlation analyses were performed between performance in EVO Monitor and SDMT. In an exploratory analysis, the correlation between performance in EVO Monitor and both physical MS measures (T25FW) and MRI volumetric data (thalamus and frontal lobes) was assessed with Pearson correlation analyses. Partial correlation analyses including age, sex, years of education, and BRT as covariates were applied when appropriate. All numerical data are presented as the mean and SEM. The statistical analyses were performed using IBM SPSS Statistics version 22.0 (IBM Corp). The significance of the statistical level was set at $P \leq .05$.

Results

Participant Characteristics

A total of 100 participants with MS and 24 participants without MS were enrolled in the study. For analysis purposes, the 100 participants with MS were divided into CI (n=33) and non-CI (n=67) subgroups. Table 1 summarizes their clinical and demographic characteristics. Figure 2 details the completion rate for each test.

Table 1. Demographic and clinical characteristics of participants.

Characteristic	MS[a]		Non-MS (n=24)
	CI[b] (n=33)	Non-CI (n=67)	
Age (years), mean (SEM)	48.96 (2.29)	53.80 (1.44)[c]	46.04 (3.72)
Sex (female), n (%)	24 (73)	51 (76)	12 (50)
Education (years), mean (SEM)	16.36 (0.41)	16.70 (0.30)	16.16 (0.41)
Handedness (right-handed), n (%)	30 (90)	59 (88)	24 (100)
Race, n (%)			
White	29 (88)	56 (84)	19 (79)
Black or African American	1 (3)	4 (6)	1 (4)
Other or unknown	3 (9)	7 (10)	4 (17)
SDMT[d] score, mean (SEM)	35.18 (6.43)	49.79 (0.98)[e]	51.20 (2.65)[f]
SDMT *z* score, mean (SEM)	–1.53 (0.06)	0.14 (0.09)[e]	0.26 (0.20)[f]
EDSS[g], median (IQR)	4 (2.75)	3 (2)	N/A[h]
Disease duration (years), mean (SEM)	11.65 (1.54)	13.27 (1.05)	N/A
MS subtype, n (%)			
Relapsing-remitting	26 (79)	48 (72)	N/A
Primary progressive	2 (6)	9 (13.5)	N/A
Secondary progressive	4 (12)	7 (10.5)	N/A
CIS[i]	0 (0)	2 (3)	N/A
Unknown	1 (3)	1 (1)	N/A

[a] MS: multiple sclerosis.

[b] CI: cognitive impairment.

[c] *P*=.04 for the comparison between non-MS and non-CI groups.

[d] SDMT: Symbol Digit Modalities Test.

[e] *P*<.001 for the comparison between CI and non-CI groups.

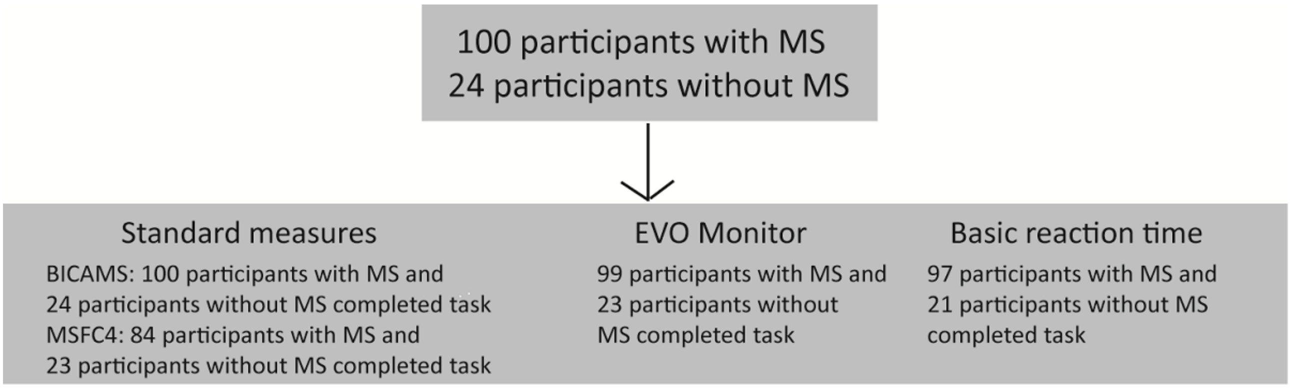
[f] *P*<.001 for the comparison between CI and non-MS groups.

[g] EDSS: Expanded Disability Status Scale.

[h] N/A: not applicable.

[i] CIS: clinically isolated syndrome.

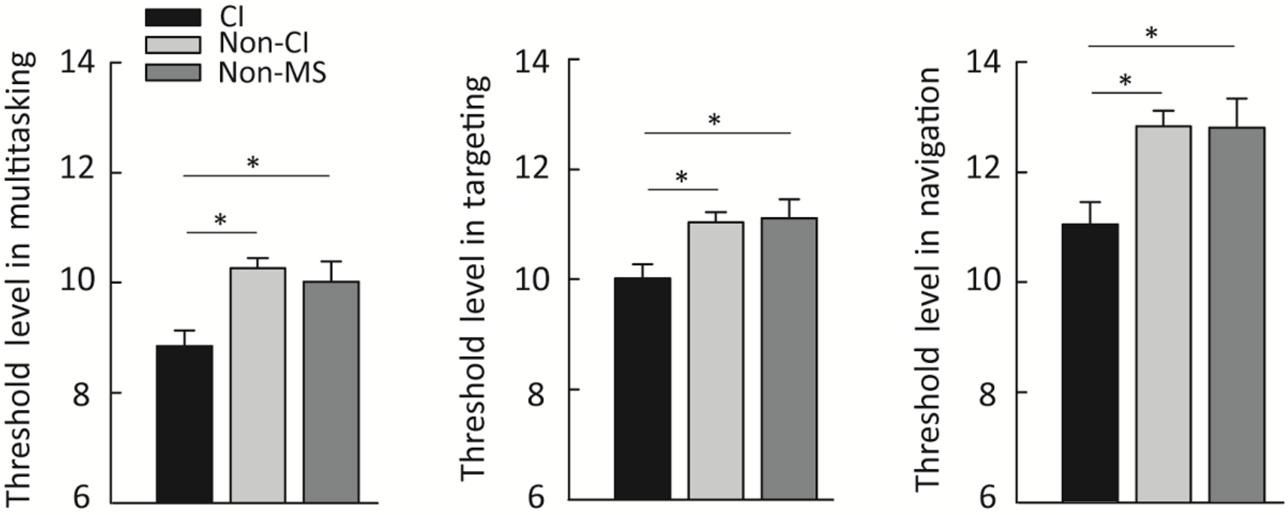
Figure 2. Study task completion rate. BICAMS: Brief International Cognitive Assessment for Multiple Sclerosis; MS: multiple sclerosis; MSFC: Multiple Sclerosis Functional Composite.



Group Differences

To evaluate whether EVO Monitor is sensitive to differences between participants with MS with and without CI as well as participants without MS, one-way ANCOVA with age, sex, years of education, and BRT as covariates was performed for the threshold level, which reflects task performance. Significant group differences in multitasking ($F_{2,109}$=8.33, *P*<.001), perceptual discrimination ($F_{2,109}$=5.63, *P*=.005), and visuomotor tracking ($F_{2,109}$=6.97, *P*=.001) threshold level were found. Post hoc analyses showed a lower threshold level in participants with CI compared with both participants without CI with MS and participants without MS in all 3 conditions (multitasking: CI vs non-CI, 8.85 [SEM 0.28] vs 10.26 [SEM 0.19]; *P*<.001; CI vs non-MS, 8.85 [SEM 0.28] vs 10.01 [SEM 0.37]; *P*=.01; perceptual discrimination: CI vs non-CI, 10.01 [SEM 0.26] vs 11.04 [SEM 0.18]; *P*=.002; CI vs non-MS, 10.01 [SEM 0.26] vs 11.11 [SEM 0.34]; *P*=.01; visuomotor tracking: CI vs non-CI, 11.05 [SEM 0.40] vs 12.83 [SEM 0.28]; *P*<.001; CI vs non-MS, 11.05 [SEM 0.40] vs 12.81 [SEM 0.52]; *P*=.01) (Figure 3). These findings indicate that EVO Monitor, a video game–based assessment designed to assess executive function, attention, and information processing speed, is sensitive to capture group-level differences between participants with MS with or without CI as well as participants without MS.

Figure 3. Group differences in EVO Monitor performance between CI, non-CI participants with MS and non-MS participants. Error bars represent standard error of the mean. CI: cognitive impairment; MS: multiple sclerosis. **P*≤.01.



Association Between Performance in EVO Monitor and Standard Measures

Pearson correlation analyses were performed to scrutinize associations between performance in EVO Monitor and standard cognitive measures. The SDMT showed significant correlations with the EVO multitasking threshold level (Figure 4 and Table 2). Including age, sex, years of education, and BRT as covariates did not change the results (Table 2). Restricting the analyses to only participants with MS showed similar results. Furthermore, associations between clinical characteristics (ie, Expanded Disability Status Scale [EDSS] and disease duration) and the EVO multitasking threshold level were observed. Analyses of EVO perceptual discrimination and visuomotor tracking threshold levels showed similar results (Table 2).

Figure 4. Correlation between EVO Monitor performance and SDMT score. SDMT: Symbol Digit Modalities Test.

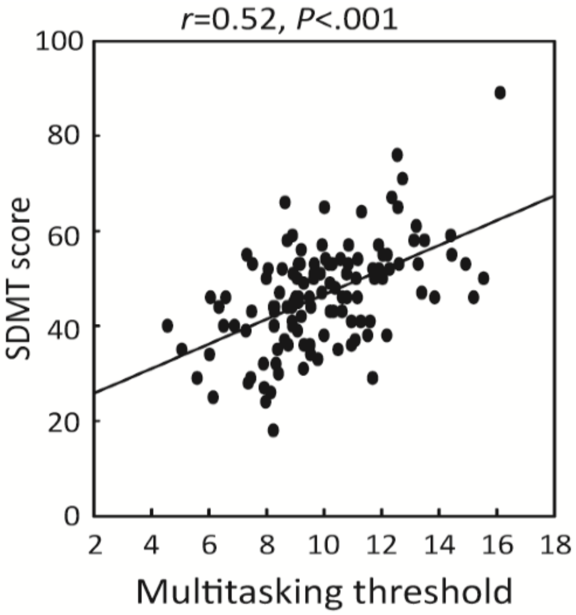


Table 2. Correlation between performance in standard cognitive measures and EVO Monitor.

EVO measures, population, and standard measures	Covariates	*r*	*P* value
Multitasking threshold level			
All participants			
SDMT[a]	N/A[b]	0.52	<.001
SDMT	Age, sex, education (years), and BRT[c]	0.38	<.001
Participants with MS[d]			
SDMT	N/A	0.46	<.001
SDMT	Age, sex, education (years), and BRT	0.41	<.001
EDSS[e]	N/A	–0.31	.002
Disease duration	N/A	–0.26	.01
Perceptual discrimination threshold level			
All participants			
SDMT	N/A	0.50	<.001
SDMT	Age, sex, education (years), and BRT	0.40	<.001
Participants with MS			
SDMT	N/A	0.39	<.001
SDMT	Age, sex, education (years), and BRT	0.34	.001
EDSS	N/A	–0.31	<.001
Disease duration	N/A	–0.29	.003
Visuomotor tracking threshold level			
All participants			
SDMT	N/A	0.49	<.001
SDMT	Age, sex, education (years), and BRT	0.38	<.001
Participants with MS			
SDMT	N/A	0.44	.01
SDMT	Age, sex, education (years), and BRT	0.40	<.001
EDSS	N/A	–0.25	.01
Disease duration	N/A	–0.21	.03

[a] SDMT: Symbol Digit Modalities Test.

[b] N/A: not applicable.

[c] BRT: basic reaction time.

[d] MS: multiple sclerosis.

[e] EDSS: Expanded Disability Status Scale.

In a planned exploratory analysis, we investigated the correlation between performance in EVO Monitor and both MS-related physical function (walking speed, T25FW) and MRI volumetric data, given that these factors have been reported to associate with cognition in MS [31-34]. A negative correlation was observed between T25FW and multitasking threshold level (*r*=–0.45, *P*<.001) (Figure 5). The results remained similar when including demographic features as covariates (Table 3). Clinically acquired MRI volumetric data were available in 56 participants with MS (17 with CI). We focused on the thalamus and frontal lobes, given their reported relationships with cognition in MS [33,34]. The analyses revealed a positive correlation between the EVO perceptual discrimination threshold level and bilateral thalamic (left: *r*=0.47, *P*<.001; right: *r*=0.39, *P*=.002) as well as rostral middle frontal (left: *r*=0.28, *P*=.03; right: *r*=0.27, *P*=.03) volumes (Figure 6). The association with thalamic volumes persisted after adjusting for age, sex, years of education, and BRT (left: *r*=0.49, *P*<.001; right: *r*=0.38, *P*=.007). These results suggest that EVO Monitor performance is associated with standard MS cognitive and physical measures as well as MRI volumetric data.

Figure 5. Correlation between EVO Monitor performance and T25FW. T25FW: Timed 25-Foot Walk.

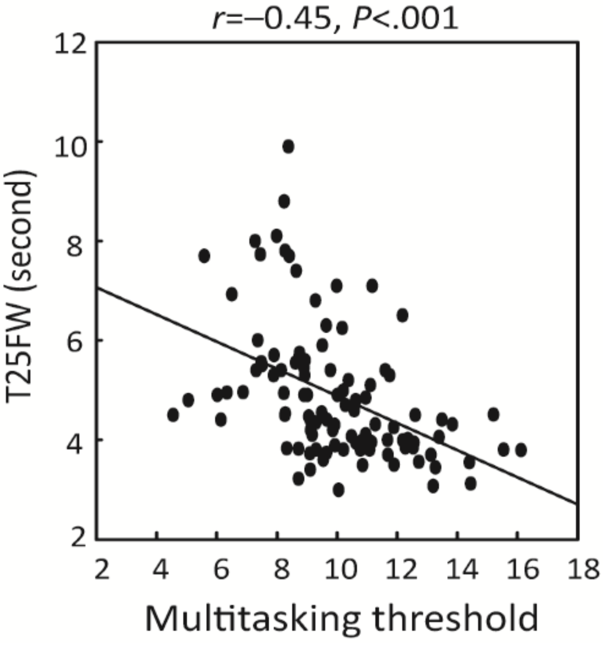


Table 3. Results of correlation between physical measure (Timed 25-Foot Walk) and EVO Monitor performance.

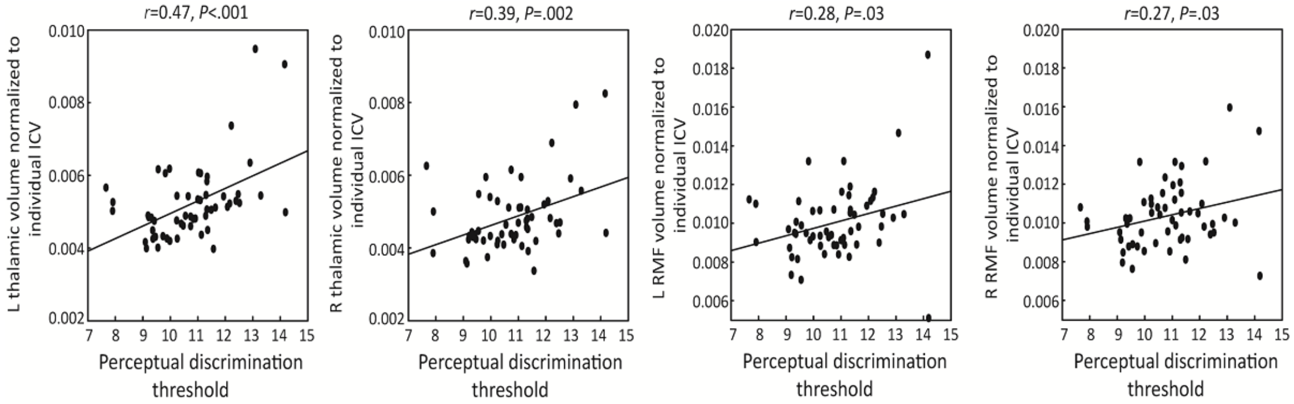
EVO measures, population, and covariates	*r*	*P* value
Multitasking threshold level		
All participants		
N/A[a]	–0.45	<.001
Age, sex, education (years), and BRT[b]	–0.30	.003
Participants with MS[c]		
N/A	–0.41	<.001
Age, sex, education (years), and BRT	–0.34	.002
Perceptual discrimination threshold level		
All participants		
N/A	–0.37	<.001
Age, sex, education (years), and BRT	–0.22	.02
Participants with MS		
N/A	–0.29	.006
Age, sex, education (years), and BRT	–0.21	.06
Visuomotor tracking threshold level		
All participants		
N/A	–0.48	<.001
Age, sex, education (years), and BRT	–0.38	<.001
Participants with MS		
N/A	–0.43	<.001
Age, sex, education (years), and BRT	–0.40	<.001

[a] N/A: not applicable.

[b] BRT: basic reaction time.

[c] MS: multiple sclerosis.

Figure 6. Correlation between EVO Monitor performance and magnetic resonance imaging volumetric data. ICV: intracranial volume; L: left; R: right; RMF: rostral middle frontal.



Discussion

In this study, we aimed to determine whether in-game assessment features of EVO Monitor, an unsupervised, digital, video game–based tool integrated with adaptive algorithms, could represent a novel and sensitive way to perform unsupervised cognitive evaluations in MS. We found a significant group-level difference in performance on EVO Monitor among participants with MS with and without CI as well as participants without MS. Furthermore, we discovered an association between performance in EVO Monitor and standard cognitive measures, physical measures, and MRI volumetric data. These results provide evidence that EVO Monitor, an unsupervised, tablet game–based digital tool designed to assess executive function, attention, and information processing, could effectively assess cognitive performance in people with MS.

We demonstrated that EVO Monitor is sensitive enough to reveal group-level differences, as shown by a lower threshold level in participants with MS with CI compared to both participants with MS without CI and participants without MS in all 3 task conditions. The EVO Monitor program was developed to challenge nearly all aspects of cognition, including executive function, attention, and information processing. In perceptual discrimination and visuomotor tracking tasks, fast information processing is highly demanded, given that the participants are asked to make a fast response to certain stimuli (ie, colored target) and rapidly changing environment (ie, dynamically moving road with walls and obstacles), respectively. To successfully perform the multitasking challenge, there are additional cognitive requirements for selective attention, sustained attention, task switching, and goal management. Since processing speed, attention control, and executive function are the most commonly affected cognitive domains in MS [4,5], the lower EVO Monitor performance in participants with MS and CI may reflect a lower performance in these cognitive domains in general. These results suggest that cognitive dysfunction in MS can be captured by EVO Monitor, a digital tool with adaptive algorithms developed for cognitive assessment.

Using unsupervised digital tools provides a less stressful context and a high level of standardization of assessment. Moreover, it can be easily applied in multiple settings, including patients' homes, which substantially improves access to these therapies for patients who may face scheduling, geographic, or economic barriers in accessing standard forms of cognitive assessment. With the advances in digital therapeutics, assessment and health care services for people with MS have been transitioning to digital platforms [25,26,52]. These results extended the application of digital tools for cognitive assessments in MS by incorporating advanced visualization and reward loops with closed-loop adaptation mechanics to enhance engagement and reduce interindividual variability, which leads to more reliable assessments [8,47].

An association between performance in EVO Monitor and standard measures was observed. Participants with a better performance in EVO Monitor showed higher SDMT scores. SDMT is considered the most sensitive measurement for the evaluation of cognitive involvement and information processing speed in the MS course [29,30]. While the participant substitutes geometric symbols for numbers while scanning a response key, a number of domains—processing speed, sustained attention, visual scanning and tracking, and motoric quickness—are being challenged. Although the task structures are different between EVO Monitor and SDMT, the shared cognitive functions subserving the two tasks may explain the correlational links. The results suggest that EVO Monitor can reliably reflect performance in cognitive tasks that involve information processing speed, selective attention, sustained attention, visual tracking, and goal management. Importantly, analyses including only participants with MS did not change the results but further revealed a negative correlation between clinical characteristics (ie, EDSS and disease duration) and performance in EVO Monitor, where participants with MS who had a longer disease duration and higher level of disability showed a lower performance in EVO Monitor. Although the correlation was modest, given that cognitive deficits tend to progress with disease duration [53] and disability progression [54] in MS, the observed association was expected.

In an exploratory analysis, we discovered an association between EVO Monitor performance and the T25FW as well as the bilateral rostral middle frontal and thalamic volume. In line with previous studies reporting that cognition in MS is associated with physical fitness and balance [55,56] and the concept of cognitive-motor coupling [57], in this study we found that better EVO Monitor performance is correlated with a faster walking speed as measured by the T25FW. Imaging [58] and postmortem [59] studies have suggested that frontal and subcortical regions involved in executive function and cognitive processing speed are also related to the spatial and temporal aspects of gait. The observed association between EVO Monitor performance and the T25FW may be explained by cerebral injury causing impairment in both domains in MS.

Additionally, the performance in the perceptual discrimination task of EVO Monitor positively correlated with rostral middle frontal and thalamic volume. The rostral middle frontal region is part of the dorsolateral prefrontal cortex, an area that is considered the center of executive function [60] and several domains of cognitive control abilities required to perform the EVO perceptual discrimination task, including sustained attention [61], selective attention [62], and inhibitory control [63]. Studies have also demonstrated that the thalamus mediates arousal states [64], and thalamic atrophy is the most significant MRI correlate of CI in MS [65,66]. It is unclear why performance in multitasking, a task including both perceptual discrimination and visuomotor tracking, did not correlate with thalamus and frontal volumetric data. Given that the posterior cortex has been linked to sustaining attention to spatial locations [67], a key component of the visuomotor task, it is possible that the multitasking task does not solely rely on frontal and thalamic resources but is also supported by the posterior cortex. Therefore, the examined correlation was not strong enough to be detected, as it was in the perceptual discrimination task.

Although our analyses were exploratory, the observed correlations support the need to better understand how EVO Monitor performance is associated with physical performance and the structural and functional changes of the brain in people with MS. Specifically, studies with functional MRI and connectivity data would provide essential information about neural correlates of the EVO Monitor tasks and, more importantly, pathological changes related to CI and neural plastic changes as a result of cognitive remediation.

Previously, we showed that EVO-AKL-T01, a video game–based digital tool similar to EVO Monitor, is an effective in-home cognitive remediation program for MS [17,18]. The high adherence rate during the 4- to 6-week home-based cognitive rehabilitation strategy indicated that remote digital tools are well accepted by patients with MS, who may have limited access to cognitive assessment or treatment. Since EVO Monitor is a digital tool designed as a self-guided assessment, it can be used as an in-home cognitive evaluation with multiple assessments to track either the progression of CI or the responsiveness to cognitive interventions, which can substantially help patients navigate problems related to cognitive issues. Future studies are warranted to evaluate the use of EVO Monitor in everyday situations across different contexts and investigate whether the results would be different due to fluctuations in cognitive ability during the day. Another key factor in cognitive assessment is patient fatigue. Well-designed studies that control for impacts of perceived fatigue and fatigability on cognitive assessments measured by digital tools are needed. Future studies comparing how different aspects of fatigue (eg, physical vs cognitive fatigue) affect cognitive performance as assessed by a comprehensive neuropsychological examination and by a digital tool that provides a shorter assessment time would provide additional insight into the role of fatigue in digital cognitive assessments.

There are some limitations to this study. Since we only included one time point in this cross-sectional analysis, it is difficult to determine the reproducibility of the observed results to conclude the test-retest reliability of EVO Monitor. In the exploratory analysis of the association between performance in EVO Monitor and MRI volumetric measures, clinically acquired MRI scans were only available for about half of the participants with MS (n=56), and the impact of acquisition protocol heterogeneity on our MRI metrics should be taken into account while a robust image processing pipeline was applied. The age difference between the non-MS group and non-CI MS group is one of the caveats in this study. However, in our analyses, age was included as a covariate to control for the potential influence of age difference on the results. Skills using digital tools may be a confounding factor influencing the results, as participants with better digital tool skills or more experience using tablet devices may have performed better. Future studies investigating digital tools should control for participants' experience and skills using digital devices.

This study extended the application of digital tools for cognitive assessments in MS by incorporating built-in adaptive staircase algorithms to enhance engagement and mitigate interindividual variability. Furthermore, the encouraging findings suggest that EVO Monitor, an unsupervised, tablet game–based program, is a clinically valuable approach to capturing CI in MS. Since CI is one of the most debilitating manifestations of MS and traveling to clinics may be burdensome due to deficits in mobility or cognition, some patients may have limited access to cognitive assessments. The application of digital cognitive assessments provides flexibility, as the testing can be performed in different settings, including patients' homes. The development of digital cognitive assessments helps patients effectively detect CI and navigate cognition-related problems in their daily living. Future studies with multiple points of data collection and a deeper investigation of how physical performance and the functional and structural changes of the brain affect cognitive performance as measured by digital tools are warranted to provide additional insight.

Acknowledgments

We would like to thank our research participants. We thank Andrew Heusser, Guillaume Poirier, Vincent Hennemand, and Denise Bonet from Akili Interactive for suggestions concerning the study design.

This work was supported by the Doris Duke Charitable Foundation and Akili Interactive. Akili Interactive provided study tools (EVO Monitor) without charge for the study.

Conflicts of Interest

RB has received research support from the National Multiple Sclerosis Society, the Hilton Foundation, the California Initiative to Advance Precision Medicine, the Sherak Foundation, and Akili Interactive. RB has also received personal compensation for consulting from Alexion, Biogen, EMD Serono, Novartis, Pear Therapeutics, Roche Genentech, and Sanofi Genzyme. RH has received personal compensation for consulting from Roche, Novartis, MEDDAY, Sanofi, Atara, and QIA. AG is cofounder, shareholder, board of directors member, and advisor for Akili Interactive Labs, a company that manufactures investigational digital treatments delivered through a video game–like interface. AG has a patent for a game-based cognitive assessment on which the tool (EVO Monitor) that was used in this study was based. All other authors declare no conflicts of interest.

References

1. Rao SM, Leo GJ, Ellington L, Nauertz T, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning. Neurology 1991 May;41(5):692-696. [doi: 10.1212/wnl.41.5.692] [Medline: 1823781]
2. Ruet A, Deloire M, Hamel D, Ouallet J, Petry K, Brochet B. Cognitive impairment, health-related quality of life and vocational status at early stages of multiple sclerosis: a 7-year longitudinal study. J Neurol 2013 Mar;260(3):776-784. [doi: 10.1007/s00415-012-6705-1] [Medline: 23081755]
3. Amato MP, Langdon D, Montalban X, Benedict RHB, DeLuca J, Krupp LB, et al. Treatment of cognitive impairment in multiple sclerosis: position paper. J Neurol 2013 Jun;260(6):1452-1468. [doi: 10.1007/s00415-012-6678-0] [Medline: 23180174]
4. Achiron A, Barak Y. Cognitive impairment in probable multiple sclerosis. J Neurol Neurosurg Psychiatry 2003 Apr;74(4):443-446 [FREE Full text] [doi: 10.1136/jnnp.74.4.443] [Medline: 12640060]
5. Rogers JM, Panegyres PK. Cognitive impairment in multiple sclerosis: evidence-based analysis and recommendations. J Clin Neurosci 2007 Oct;14(10):919-927. [doi: 10.1016/j.jocn.2007.02.006] [Medline: 17659875]
6. Bobholz JA, Rao SM. Cognitive dysfunction in multiple sclerosis: a review of recent developments. Curr Opin Neurol 2003;16(3):283-288. [doi: 10.1097/00019052-200306000-00006]
7. Langdon DW. Cognition in multiple sclerosis. Curr Opin Neurol 2011 Jun;24(3):244-249. [doi: 10.1097/WCO.0b013e328346a43b] [Medline: 21519256]
8. Anguera JA, Brandes-Aitken AN, Rolle CE, Skinner SN, Desai SS, Bower JD, et al. Characterizing cognitive control abilities in children with 16p11.2 deletion using adaptive 'video game' technology: a pilot study. Transl Psychiatry 2016 Sep 20;6(9):e893 [FREE Full text] [doi: 10.1038/tp.2016.178] [Medline: 27648915]
9. Anguera JA, Brandes-Aitken AN, Antovich AD, Rolle CE, Desai SS, Marco EJ. A pilot study to determine the feasibility of enhancing cognitive abilities in children with sensory processing dysfunction. PLoS One 2017;12(4):e0172616 [FREE Full text] [doi: 10.1371/journal.pone.0172616] [Medline: 28380008]
10. Anguera JA, Gunning FM, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. Depress Anxiety 2017 Jun;34(6):508-517 [FREE Full text] [doi: 10.1002/da.22588] [Medline: 28052513]
11. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
12. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
13. Areàn PA, Hoa Ly K, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Jun;18(2):163-169 [FREE Full text] [Medline: 27489456]
14. Charvet LE, Yang J, Shaw MT, Sherman K, Haider L, Xu J, et al. Cognitive function in multiple sclerosis improves with telerehabilitation: Results from a randomized controlled trial. PLoS One 2017;12(5):e0177177 [FREE Full text] [doi: 10.1371/journal.pone.0177177] [Medline: 28493924]
15. Davis NO, Bower J, Kollins SH. Proof-of-concept study of an at-home, engaging, digital intervention for pediatric ADHD. PLoS One 2018;13(1):e0189749 [FREE Full text] [doi: 10.1371/journal.pone.0189749] [Medline: 29324745]
16. Flynn RM, Colón-Acosta N, Zhou J, Bower J. A Game-Based Repeated Assessment for Cognitive Monitoring: Initial Usability and Adherence Study in a Summer Camp Setting. J Autism Dev Disord 2019 May;49(5):2003-2014. [doi: 10.1007/s10803-019-03881-w] [Medline: 30656527]
17. Bove R, Rowles W, Zhao C, Anderson A, Friedman S, Langdon D, et al. A novel in-home digital treatment to improve processing speed in people with multiple sclerosis: A pilot study. Mult Scler 2020 Jun 25:1352458520930371. [doi: 10.1177/1352458520930371] [Medline: 32584155]
18. Bove RM, Rush G, Zhao C, Rowles W, Garcha P, Morrissey J, et al. A Videogame-Based Digital Therapeutic to Improve Processing Speed in People with Multiple Sclerosis: A Feasibility Study. Neurol Ther 2019 Jun;8(1):135-145 [FREE Full text] [doi: 10.1007/s40120-018-0121-0] [Medline: 30506301]

19. Lee G. Effects of training using video games on the muscle strength, muscle tone, and activities of daily living of chronic stroke patients. J Phys Ther Sci 2013 May;25(5):595-597 [FREE Full text] [doi: 10.1589/jpts.25.595] [Medline: 24259810]
20. Webster D, Celik O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. J Neuroeng Rehabil 2014;11:108 [FREE Full text] [doi: 10.1186/1743-0003-11-108] [Medline: 24996956]
21. Trapp W, Landgrebe M, Hoesl K, Lautenbacher S, Gallhofer B, Günther W, et al. Cognitive remediation improves cognition and good cognitive performance increases time to relapse – results of a 5 year catamnestic study in schizophrenia patients. BMC Psychiatry 2013 Jul 9;13(1):184. [doi: 10.1186/1471-244x-13-184]
22. Kollins SH, Bower J, Findling RL, Keefe R, Epstein J, Cutler AJ, et al. 2.40 A Multicenter, Randomized, Active-Control Registration Trial of Software Treatment for Actively Reducing Severity of ADHD (Stars-Adhd) to Assess the Efficacy and Safety of a Novel, Home-Based, Digital Treatment for Pediatric ADHD. J Am Acad Child Adolesc Psychiatry 2018 Oct;57(10):S172. [doi: 10.1016/j.jaac.2018.09.128]
23. Kollins S, DeLoss D, Cañadas E, Lutz J, Findling R, Keefe R, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digital Health 2020 Apr;2(4):e168-e178 [FREE Full text] [doi: 10.1016/s2589-7500(20)30017-0]
24. Yerys BE, Bertollo JR, Kenworthy L, Dawson G, Marco EJ, Schultz RT, et al. Brief Report: Pilot Study of a Novel Interactive Digital Treatment to Improve Cognitive Control in Children with Autism Spectrum Disorder and Co-occurring ADHD Symptoms. J Autism Dev Disord 2019 Apr;49(4):1727-1737. [doi: 10.1007/s10803-018-3856-7] [Medline: 30569407]
25. Finkelstein J, Lapshin O, Castro H, Cha E, Provance PG. Home-based physical telerehabilitation in patients with multiple sclerosis: a pilot study. J Rehabil Res Dev 2008;45(9):1361-1373 [FREE Full text] [Medline: 19319760]
26. Hatzakis M, Haselkorn J, Williams R, Turner A, Nichol P. Telemedicine and the delivery of health services to veterans with multiple sclerosis. J Rehabil Res Dev 2003;40(3):265-282 [FREE Full text] [Medline: 14582530]
27. Kane RL, Bever CT, Ehrmantraut M, Forte A, Culpepper WJ, Wallin MT. Teleneurology in patients with multiple sclerosis: EDSS ratings derived remotely and from hands-on examination. J Telemed Telecare 2008;14(4):190-194. [doi: 10.1258/jtt.2008.070904] [Medline: 18534953]
28. Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowich J, et al. Video game training enhances cognitive control in older adults. Nature 2013 Sep 5;501(7465):97-101 [FREE Full text] [doi: 10.1038/nature12486] [Medline: 24005416]
29. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, Multiple Sclerosis Outcome Assessments Consortium. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult Scler 2017 Apr;23(5):721-733 [FREE Full text] [doi: 10.1177/1352458517690821] [Medline: 28206827]
30. Parmenter BA, Weinstock-Guttman B, Garg N, Munschauer F, Benedict RHB. Screening for cognitive impairment in multiple sclerosis using the Symbol digit Modalities Test. Mult Scler 2007 Jan;13(1):52-57. [doi: 10.1177/1352458506070750] [Medline: 17294611]
31. Sandroff BM, Hillman CH, Motl RW. Aerobic fitness is associated with inhibitory control in persons with multiple sclerosis. Arch Clin Neuropsychol 2015 Jun;30(4):329-340. [doi: 10.1093/arclin/acv022] [Medline: 25910783]
32. Sandroff BM, Pilutti LA, Benedict RHB, Motl RW. Association between physical fitness and cognitive function in multiple sclerosis: does disability status matter? Neurorehabil Neural Repair 2015;29(3):214-223. [doi: 10.1177/1545968314541331] [Medline: 25009224]
33. Rocca MA, Amato MP, De Stefano N, Enzinger C, Geurts JJ, Penner I, MAGNIMS Study Group. Clinical and imaging assessment of cognitive dysfunction in multiple sclerosis. Lancet Neurol 2015 Mar;14(3):302-317. [doi: 10.1016/S1474-4422(14)70250-9] [Medline: 25662900]
34. Rocca MA, Riccitelli GC, Meani A, Pagani E, Del Sette P, Martinelli V, et al. Cognitive reserve, cognition, and regional brain damage in MS: A 2 -year longitudinal study. Mult Scler 2019 Mar;25(3):372-381. [doi: 10.1177/1352458517750767] [Medline: 29303036]
35. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011 Feb;69(2):292-302 [FREE Full text] [doi: 10.1002/ana.22366] [Medline: 21387374]
36. Corfield F, Langdon D. A Systematic Review and Meta-Analysis of the Brief Cognitive Assessment for Multiple Sclerosis (BICAMS). Neurol Ther 2018 Dec;7(2):287-306 [FREE Full text] [doi: 10.1007/s40120-018-0102-3] [Medline: 29923070]
37. Langdon D, Amato M, Boringa J, Brochet B, Foley F, Fredrikson S, et al. Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). Mult Scler 2012 Jun;18(6):891-898 [FREE Full text] [doi: 10.1177/1352458511431076] [Medline: 22190573]
38. Delis D, Kramer J, Kaplan E, Ober B. California verbal learning test-II. 2nd ed. San Antonio, TX: The Psychological Corporation; 2000.
39. Benedict RH. Brief visuospatial memory test - revised: professional manual. Lutz, FL: Psychological Assessment Resources Inc; 1997.
40. Benedict RHB, Duquin JA, Jurgensen S, Rudick RA, Feitcher J, Munschauer FE, et al. Repeated assessment of neuropsychological deficits in multiple sclerosis using the Symbol Digit Modalities Test and the MS Neuropsychological Screening Questionnaire. Mult Scler 2008 Aug;14(7):940-946. [doi: 10.1177/1352458508090923] [Medline: 18573822]

41. Cohen JA, Reingold SC, Polman CH, Wolinsky JS. Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects. The Lancet Neurology 2012 May;11(5):467-476. [doi: 10.1016/S1474-4422(12)70059-5]
42. Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. Percept Mot Skills 1977 Apr;44(2):367-373. [doi: 10.2466/pms.1977.44.2.367] [Medline: 866038]
43. García-Pérez MA. Adaptive psychophysical methods for nonmonotonic psychometric functions. Atten Percept Psychophys 2014 Feb;76(2):621-641. [doi: 10.3758/s13414-013-0574-2] [Medline: 24197504]
44. King-Smith PE, Rose D. Principles of an adaptive method for measuring the slope of the psychometric function. Vision Res 1997 Jun;37(12):1595-1604 [FREE Full text] [doi: 10.1016/s0042-6989(96)00310-0] [Medline: 9231226]
45. Leek MR. Adaptive procedures in psychophysical research. Percept Psychophys 2001 Nov;63(8):1279-1292. [doi: 10.3758/bf03194543] [Medline: 11800457]
46. Klein SA. Measuring, estimating, and understanding the psychometric function: a commentary. Percept Psychophys 2001 Nov;63(8):1421-1455. [doi: 10.3758/bf03194552] [Medline: 11800466]
47. Mishra J, Anguera JA, Gazzaley A. Video Games for Neuro-Cognitive Optimization. Neuron 2016 Dec 20;90(2):214-218 [FREE Full text] [doi: 10.1016/j.neuron.2016.04.010] [Medline: 27100194]
48. Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 2014 Oct 01;99:166-179. [doi: 10.1016/j.neuroimage.2014.05.044] [Medline: 24879923]
49. Klein A, Ghosh SS, Bao FS, Giard J, Häme Y, Stavsky E, et al. Mindboggling morphometry of human brains. PLoS Comput Biol 2017 Feb;13(2):e1005350 [FREE Full text] [doi: 10.1371/journal.pcbi.1005350] [Medline: 28231282]
50. Whitwell JL, Crum WR, Watt HC, Fox NC. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. AJNR Am J Neuroradiol 2001 Sep;22(8):1483-1489 [FREE Full text] [Medline: 11559495]
51. Kiely KM, Butterworth P, Watson N, Wooden M. The Symbol Digit Modalities Test: Normative data from a large nationally representative sample of Australians. Arch Clin Neuropsychol 2014 Dec;29(8):767-775. [doi: 10.1093/arclin/acu055] [Medline: 25352087]
52. Hoang P, Schoene D, Gandevia S, Smith S, Lord SR. Effects of a home-based step training programme on balance, stepping, cognition and functional performance in people with multiple sclerosis--a randomized controlled trial. Mult Scler 2016 Jan;22(1):94-103. [doi: 10.1177/1352458515579442] [Medline: 25921035]
53. Amato MP, Ponziani G, Siracusa G, Sorbi S. Cognitive dysfunction in early-onset multiple sclerosis: a reappraisal after 10 years. Arch Neurol 2001 Oct;58(10):1602-1606. [doi: 10.1001/archneur.58.10.1602] [Medline: 11594918]
54. Carotenuto A, Moccia M, Costabile T, Signoriello E, Paolicelli D, Simone M, Cogniped study group. Associations between cognitive impairment at onset and disability accrual in young people with multiple sclerosis. Sci Rep 2019 Dec 02;9(1):18074 [FREE Full text] [doi: 10.1038/s41598-019-54153-7] [Medline: 31792347]
55. Batista S, Teter B, Sequeira K, Josyula S, Hoogs M, Ramanathan M, et al. Cognitive impairment is associated with reduced bone mass in multiple sclerosis. Mult Scler 2012 Oct;18(10):1459-1465. [doi: 10.1177/1352458512440206] [Medline: 22419674]
56. Sandroff BM, Motl RW. Fitness and cognitive processing speed in persons with multiple sclerosis: a cross-sectional investigation. J Clin Exp Neuropsychol 2012;34(10):1041-1052. [doi: 10.1080/13803395.2012.715144] [Medline: 22905722]
57. Benedict RHB, Holtzer R, Motl RW, Foley FW, Kaur S, Hojnacki D, et al. Upper and lower extremity motor function and cognitive impairment in multiple sclerosis. J Int Neuropsychol Soc 2011 Jul;17(4):643-653. [doi: 10.1017/S1355617711000403] [Medline: 21486517]
58. Rosano C, Aizenstein HJ, Wu M, Newman AB, Becker JT, Lopez OL, et al. Focal atrophy and cerebrovascular disease increase dementia risk among cognitively normal older adults. J Neuroimaging 2007 Apr;17(2):148-155. [doi: 10.1111/j.1552-6569.2007.00093.x] [Medline: 17441836]
59. Whitman GT, Tang Y, Lin A, Baloh RW, Tang T. A prospective study of cerebral white matter abnormalities in older people with gait dysfunction. Neurology 2001 Sep 25;57(6):990-994. [doi: 10.1212/wnl.57.6.990] [Medline: 11571322]
60. Alvarez JA, Emory E. Executive function and the frontal lobes: a meta-analytic review. Neuropsychol Rev 2006 Mar;16(1):17-42. [doi: 10.1007/s11065-006-9002-x] [Medline: 16794878]
61. Ortuño F, Ojeda N, Arbizu J, López P, Martí-Climent JM, Peñuelas I, et al. Sustained attention in a counting task: normal performance and functional neuroanatomy. Neuroimage 2002 Sep;17(1):411-420. [doi: 10.1006/nimg.2002.1168] [Medline: 12482094]
62. Hadland KA, Rushworth MF, Passingham RE, Jahanshahi M, Rothwell JC. Interference with performance of a response selection task that has no working memory component: an rTMS comparison of the dorsolateral prefrontal and medial frontal cortex. J Cogn Neurosci 2001 Nov 15;13(8):1097-1108. [doi: 10.1162/089892901753294392] [Medline: 11784448]
63. Hoppenbrouwers SS, De Jesus DR, Stirpe T, Fitzgerald PB, Voineskos AN, Schutter DJLG, et al. Inhibitory deficits in the dorsolateral prefrontal cortex in psychopathic offenders. Cortex 2013 May;49(5):1377-1385. [doi: 10.1016/j.cortex.2012.06.003] [Medline: 22795183]

64. Van der Werf YD, Witter MP, Groenewegen HJ. The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. Brain Res Brain Res Rev 2002 Sep;39(2-3):107-140. [doi: 10.1016/s0165-0173(02)00181-9] [Medline: 12423763]
65. Batista S, Zivadinov R, Hoogs M, Bergsland N, Heininen-Brown M, Dwyer MG, et al. Basal ganglia, thalamus and neocortical atrophy predicting slowed cognitive processing in multiple sclerosis. J Neurol 2012 Jan;259(1):139-146. [doi: 10.1007/s00415-011-6147-1] [Medline: 21720932]
66. Benedict RHB, Bruce JM, Dwyer MG, Abdelrahman N, Hussein S, Weinstock-Guttman B, et al. Neocortical atrophy, third ventricular width, and cognitive dysfunction in multiple sclerosis. Arch Neurol 2006 Sep;63(9):1301-1306. [doi: 10.1001/archneur.63.9.1301] [Medline: 16966509]
67. Malhotra P, Coulthard EJ, Husain M. Role of right posterior parietal cortex in maintaining attention to spatial locations over time. Brain 2009 Mar;132(Pt 3):645-660 [FREE Full text] [doi: 10.1093/brain/awn350] [Medline: 19158107]

Abbreviations

ANCOVA: analysis of covariance
BRT: basic reaction time
CI: cognitive impairment
EDSS: Expanded Disability Status Scale
MRI: magnetic resonance imaging
MS: multiple sclerosis
SDMT: Symbol Digit Modalities Test
SEM: standard error of the mean
T25FW: Timed 25-Foot Walk
UCSF: University of California, San Francisco

Edited by G Eysenbach; submitted 15.09.20; peer-reviewed by A Wright, O Rivera; comments to author 08.10.20; revised version received 14.10.20; accepted 28.10.20; published 20.01.21.

Please cite as:
Hsu WY, Rowles W, Anguera JA, Zhao C, Anderson A, Alexander A, Sacco S, Henry R, Gazzaley A, Bove R
Application of an Adaptive, Digital, Game-Based Approach for Cognitive Assessment in Multiple Sclerosis: Observational Study
J Med Internet Res 2021;23(1):e24356
URL: http://www.jmir.org/2021/1/e24356/
doi:10.2196/24356
PMID:33470940

Original Paper

Development and Implementation of a Decision Support System to Improve Control of Hypertension and Diabetes in a Resource-Constrained Area in Brazil: Mixed Methods Study

Milena Soriano Marcolino[1,2], MD, MSc, PhD; João Antonio Queiroz Oliveira[1,2], PharmD, MSc; Christiane Corrêa Rodrigues Cimini[3], MD, MSc; Junia Xavier Maia[2], MD; Vânia Soares Oliveira Almeida Pinto[3], MD, MSc; Thábata Queiroz Vivas Sá[2], MSc; Kaique Amancio[1,2]; Lissandra Coelho[3]; Leonardo Bonisson Ribeiro[2], BBA; Clareci Silva Cardoso[4], MSc, PhD; Antonio Luiz Ribeiro[1,2], MD, PhD

[1]Medical School, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[2]Telehealth Center, University Hospital, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[3]Medical School, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Teófilo Otoni, Brazil

[4]Medical School, Universidade Federal de São João del-Rei, Divinópolis, Brazil

Corresponding Author:
Milena Soriano Marcolino, MD, MSc, PhD
Medical School
Universidade Federal de Minas Gerais
Avenida Professor Alfredo Balena, 190, Santa Efigênia
Belo Horizonte, 30130100
Brazil
Phone: 55 31 98813 0688
Email: milenamarc@gmail.com

Abstract

Background: The low levels of control of hypertension and diabetes mellitus are a challenge that requires innovative strategies to surpass barriers of low sources, distance, and quality of health care.

Objective: The aim of this study is to develop a clinical decision support system (CDSS) for diabetes and hypertension management in primary care, to implement it in a resource-constrained region, and to evaluate its usability and health care practitioner satisfaction.

Methods: This mixed methods study is a substudy of HealthRise Brazil Project, a multinational study designed to implement pilot programs to improve screening, diagnosis, management, and control of hypertension and diabetes among underserved communities. Following the identification of gaps in usual care, a team of clinicians established the software functional requirements. Recommendations from evidence-based guidelines were reviewed and organized into a decision algorithm, which bases the CDSS reminders and suggestions. Following pretesting and expert panel assessment, pilot testing was conducted in a quasi-experimental study, which included 34 primary care units of 10 municipalities in a resource-constrained area in Brazil. A Likert-scale questionnaire evaluating perceived feasibility, usability, and utility of the application and professionals' satisfaction was applied after 6 months. In the end-line assessment, 2 focus groups with primary care physicians and nurses were performed.

Results: A total of 159 reminders and suggestions were created and implemented for the CDSS. At the 6-month assessment, there were 1939 patients registered in the application database and 2160 consultations were performed by primary care teams. Of the 96 health care professionals who were invited for the usability assessment, 26% (25/96) were physicians, 46% (44/96) were nurses, and 28% (27/96) were other health professionals. The questionnaire included 24 items on impressions of feasibility, usability, utility, and satisfaction, and presented global Cronbach α of .93. As for feasibility, all professionals agreed (median scores of 4 or 5) that the application could be used in primary care settings and it could be easily incorporated in work routines, but physicians claimed that the application might have caused significant delays in daily routines. As for usability, overall evaluation was good and it was claimed that the application was easy to understand and use. All professionals agreed that the application was useful (score 4 or 5) to promote prevention, assist treatment, and might improve patient care, and they were overall satisfied with the application (median scores between 4 and 5). In the end-line assessment, there were 4211 patients (94.82% [3993/4211] with hypertension and 24.41% [1028/4211] with diabetes) registered in the application's database and 7960 consultations were performed by primary health care teams. The 17 participants of the focus groups were consistent to affirm they were very satisfied with the CDSS.

Conclusions: The CDSS was applicable in the context of primary health care settings in low-income regions, with good user satisfaction and potential to improve adherence to evidence-based practices.

(J Med Internet Res 2021;23(1):e18872) doi:10.2196/18872

KEYWORDS

clinical decision support systems; primary health care; hypertension; diabetes mellitus; evidence-based practice; telemedicine; patient care management

Introduction

Hypertension and diabetes are leading modifiable risk factors for cardiovascular disease worldwide, major contributors to premature disability, and associated with substantial premature death and morbidity [1]. Despite all the advances in the therapy for these diseases, and the fact that many effective treatments are available, there is a great deal of room for improvement. There is a large gap between detection and control of both diseases, and the majority of patients do not reach therapeutic goals [2-4]. The situation is worse in low- and middle-income countries and even worse in rural communities, where awareness, treatment, and control of hypertension are lower than in urban areas [2,4,5].

To face this challenge, innovative strategies are required. Clinical decision support systems (CDSSs), which are capable of generating suggestions or information that is specific to individual patients, based on the unique individualized patient information, can make a significant contribution to the effective dissemination of evidence-based practice. They may increase practitioners' adherence to guidelines, involving clinicians in the translation of research into practice [6]. Therefore, they may increase the proportion of patients who reach blood pressure and glucose goals, with a potential impact in reducing cardiovascular risk.

However, the evidence of impact of CDSSs on key diabetes care outcomes, such as the control of glucose, blood pressure, tobacco use, or appropriate aspirin use has been marginal or inconsistent. One important barrier in this context is CDSS usability [7]. Many CDSSs were not used regularly or on a sustained basis by primary care physicians [8]. This is true not only for diabetes, but also for CDSS in general. It has been reported that their adoption has been somewhat limited outside of a relatively small number of academic medical centers and integrated health care delivery networks [9,10]. Therefore, in order to translate evidence into practice and develop CDSSs, which can be useful and used in clinical practice, studies that assess usability prior to a large-scale implementation are of utmost importance [11].

Thus, our aim was to develop a CDSS for diabetes and hypertension management in primary care, to implement it in a low-income and mostly rural region, and to evaluate its usability and health care practitioner satisfaction.

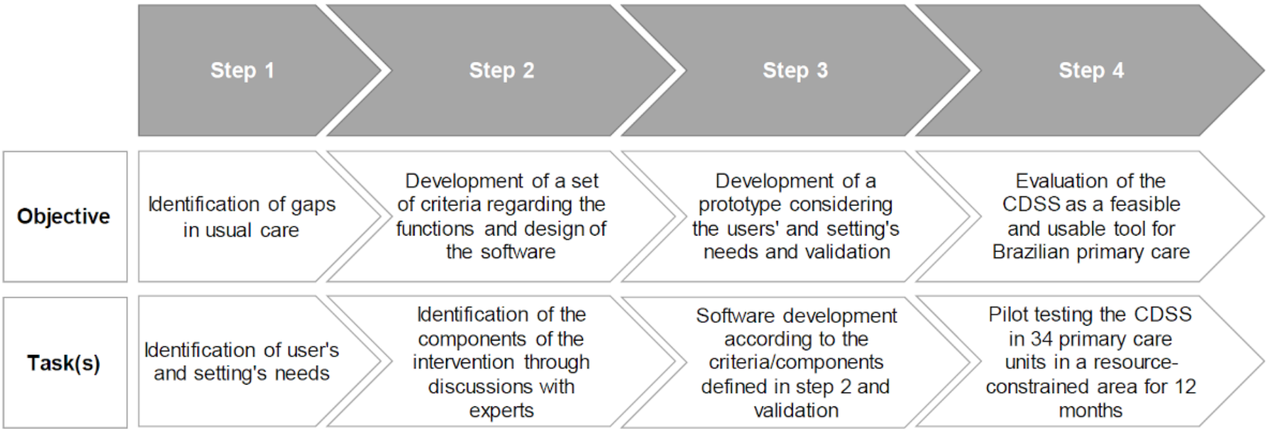
Methods

Study Overview

This is a substudy of the HealthRise Brazil Project, a quasi-experimental study, which is part of a multinational study designed to implement and evaluate pilot programs aimed at improving screening, diagnosis, management, and control of hypertension and diabetes among underserved communities [12]. Interventions were designed and implemented in 9 communities in Brazil, India, South Africa, and the United States of America. Each program was designed and implemented by a local grantee and included interventions that were tailored to local needs and contexts [12]. HealthRise Brazil Project was conducted in 2 centers, Vitória da Conquista, in the state of Bahia, and in the region of Teófilo Otoni (Vale do Mucuri), the state of Minas Gerais. In this region, the project was implemented in 10 municipalities: Frei Gaspar, Ouro Verde de Minas, Crisólita, Catuji, Setubinha, Itaipé, Novo Oriente de Minas, Ladainha, Teófilo Otoni, and Ataleia.

This study was conducted by the HealthRise Brazil Teófilo Otoni team in 4 steps, according to the Medical Research Council framework (Figure 1) [13]: (1) identification of gaps in usual care; (2) identification of the components of the intervention through discussions with experts; (3) CDSS development and validation; and (4) pilot testing. Theory of change thinking was used throughout the study: programme design, monitoring, and evaluation [14,15].

Figure 1. Flowchart of the study methodology. CDSS: clinical decision support system.



	Step 1	Step 2	Step 3	Step 4
Objective	Identification of gaps in usual care	Development of a set of criteria regarding the functions and design of the software	Development of a prototype considering the users' and setting's needs and validation	Evaluation of the CDSS as a feasible and usable tool for Brazilian primary care
Task(s)	Identification of user's and setting's needs	Identification of the components of the intervention through discussions with experts	Software development according to the criteria/components defined in step 2 and validation	Pilot testing the CDSS in 34 primary care units in a resource-constrained area for 12 months

Identification of Gaps in Usual Care

To evaluate which gaps were already known, we assessed: (1) epidemiological studies and systematic reviews on hypertension and diabetes management; (2) secondary data, including census, *Sistema de Informação Ambulatorial* (Outpatient Information System), *Sistema de Informação da Atenção Básica* (Primary Care Information System), *Cadastro Nacional de Estabelecimentos de Saúde* (National Registry of Health Establishments), *Programa Nacional de Melhoria de Acesso e da Qualidade da Atenção Básica* (National Program for Improving Access and Quality of Primary Care), social dimensions of the research "Inequalities," and *Pesquisa Nacional de Saúde* (National Health Survey). We also assessed information from workshops and focus groups performed by the sponsor, with the participation of municipal and state government health managers, physicians, pharmacists, nurses, nutritionists, physical educators, university professors, specialists, representatives of patients and local communities, community health workers (CHWs), who discussed barriers and access opportunities, as well as priority areas (diagnosis, treatment, and disease management). This information was available the moment we applied for the research funding [12], and it was the starting point to generate ideas for the intervention design [16].

Identifying the Components of the Intervention Through Discussion With Experts and Stakeholders

In order to identify the components of the intervention, information was derived from the previous step, as well as previous assessments from our group [17,18]; discussions with primary care physicians, cardiologists, endocrinologists, nurses, nutritionists, pharmacists, and physical educators. Besides, meetings and internal workshops were conducted to discuss the topic with primary care practitioners, primary care unit coordinators, local health authorities, and stakeholders. The key stakeholders were identified and involved in the project from the beginning of design of the intervention [15].

A primary care physician, an internal medicine specialist, a cardiologist, and an endocrinologist discussed the gaps with experts in information technology and stakeholders to identify solutions and components of the intervention, to map anticipated change processes, long-term changes that needed to happen in the target group's lives, the barriers to those changes, and to explore assumptions and hypotheses [16].

At that time, all primary care units used a paper-based system to manage patient records; there were different levels of internet connectivity in the primary care centers and health care workers had low technical literacy. All of these factors were taken into account.

Clinical Decision Support System Development and Validation

Software Requirements

A team of clinicians established the functional software requirements necessary to record, track, and support decision making for patients with hypertension, diabetes, or both, according to the information obtained from the previous step and also from national and international guidelines [19-29] to define the technical specifications for digital systems. A list of indicators was defined for data monitoring.

Development of Content and Functionalities

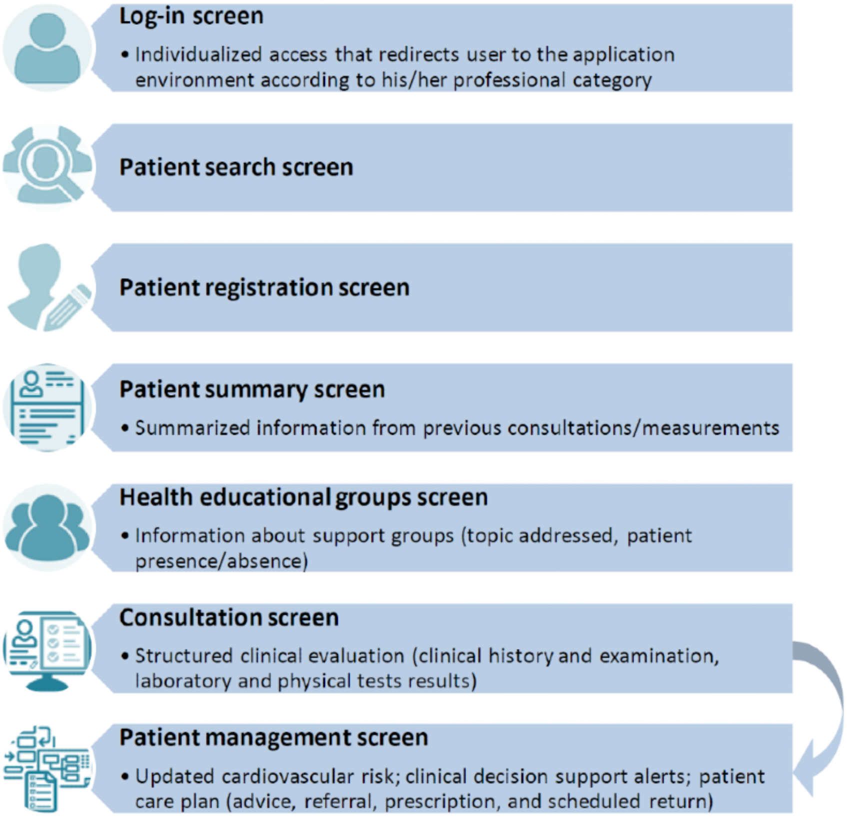
The technologies used for software development were chosen according to the available technological infrastructure in each municipality. In this context, the following aspects were taken into account: (1) existence of internet connection; (2) type of internet connection; (3) quality of internet connection; (4) existence of mobile and landline telephony; (5) equipment available in health units. Given the different realities and often precarious technology and information, the challenges faced were huge.

For the execution of the project, there was an agreement between the mayor of each municipality, the coordinator of each health unit, and the project management team to guarantee a minimum internet connection and the existence and maintenance of a local server network that would guarantee the registration and access to data while assisting patients. Despite this agreement, and because certain locations had few resources and limited or unstable internet connections, the server system was developed to work locally without the need for an internet connection. The internet was required only to synchronize local servers to the central database, whenever there was a connection, for scientific and monitoring purposes. Therefore, small drops in connection would not lead to any data loss or force an abandonment of using the system. With regard to problems in connectivity and transmission, in cases of momentary loss of synchronization of local data with the centralized data pool, data were synchronized whenever the connection was re-established. In remote areas, we used a portable internet router for transmission, and in rare instances of total absence of internet, which impaired the synchronization of the monthly data monitoring, data were copied to a portable storage device. With data from all units transmitted to the central database, these were extracted and exported to an Excel spreadsheet (Microsoft) for analysis by the research team (Multimedia Appendix 1).

The application was developed with an "user-centered design" [15], in Java language 1.8 and JSF 2.2, using NetBeans 8.1 development environment, PostgreSQL database server, and Hibernate framework, to be used in web browsers. The interface was developed to be intuitive and self-explanatory, using Prime Faces, Bootstrap, JavaScript, and jQuery. Security and access control were based on Spring Security, ensuring inviolability of the data. There was an iterative process of development, with close and daily cooperation between clinicians and developers of the application [15].

The application was designed to be used by health care professionals, and consists of: (1) a log-in screen; (2) a patient search screen; (3) a patient registration screen; (4) structured clinical evaluation, clinical decision support, and patient care plan; (5) a health educational groups screen; (6) patient summary screen; and (7) patient management screen (Figure 2).

Figure 2. The main functionalities of the application.



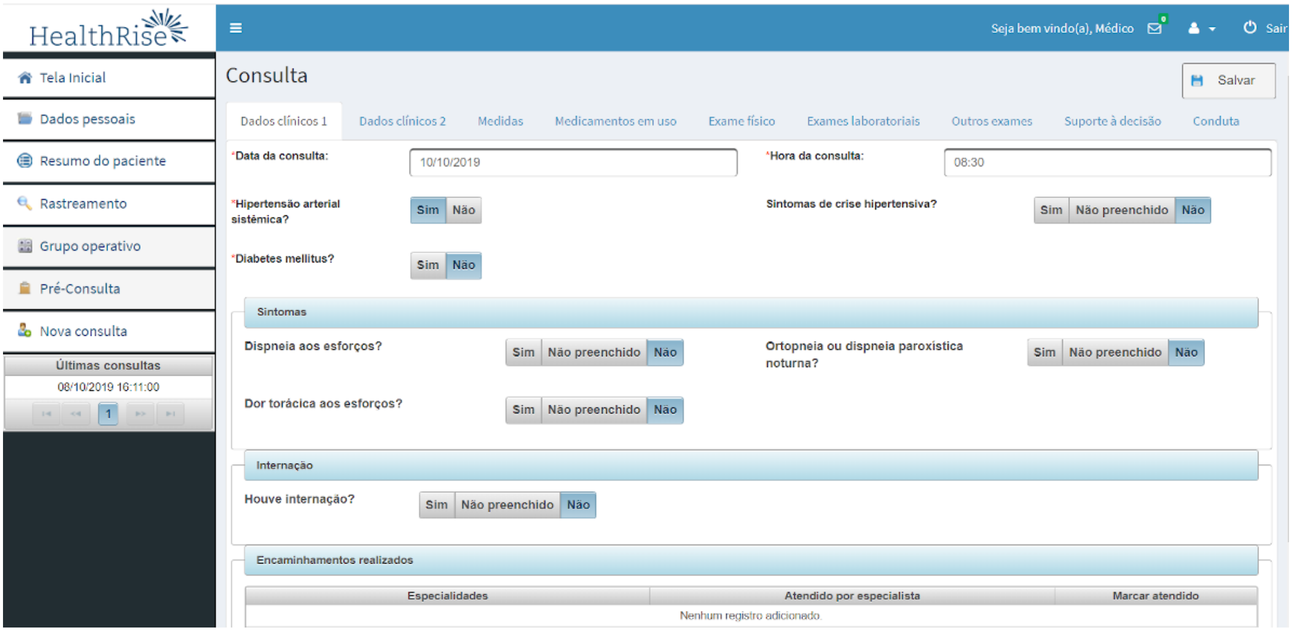
The log-in screen allows individualized access to the system through the credentials (user and password) provided to each professional. Each user is registered by a central administrator and each user account is associated with a professional profile that describes the way the application environment looks and operates for that user: physician, nurse, nurse technician, multidisciplinary primary care support team (*Núcleo Ampliado de Saúde da Família e Atenção Básica [*NASF-AB*]*, explained further in this article), and CHW.

After login, the professional has access to the screen to search for registered patients. If the patient was not registered before, he/she can access the registration screen, to register a new patient (demographic data, address, and telephone number). After registration, the other functionalities become available. Each patient has a specific identification code in the database.

The main functionalities of the application consist of the structured clinical evaluation, the clinical decision support, and the patient care plan. Data requested were manually entered, but data previously entered were saved for the next consultation, for example, medications used. In this case, the health professional could edit the information as necessary. Although few fields are mandatory, the application alerts the health care professional to the importance of completing the information appropriately and completely.

The structured clinical evaluation is shown partially in Figure 3. It includes data on symptoms, medical history, physical examination (including foot examination in patients with diabetes), current medications, and complementary examination results (laboratory and other tests).

Figure 3. Structured clinical evaluation.



*Data da consulta: 10/10/2019

*Hora da consulta: 08:30

*Hipertensão arterial sistêmica? [✓ Sim] Não

Sintomas de crise hipertensiva? Sim Não preenchido [✓ Não]

*Diabetes mellitus? Sim [✓ Não]

Dispneia aos esforços? Sim Não preenchido [✓ Não]

Ortopneia ou dispneia paroxística noturna? Sim Não preenchido [✓ Não]

Dor torácica aos esforços? Sim Não preenchido [✓ Não]

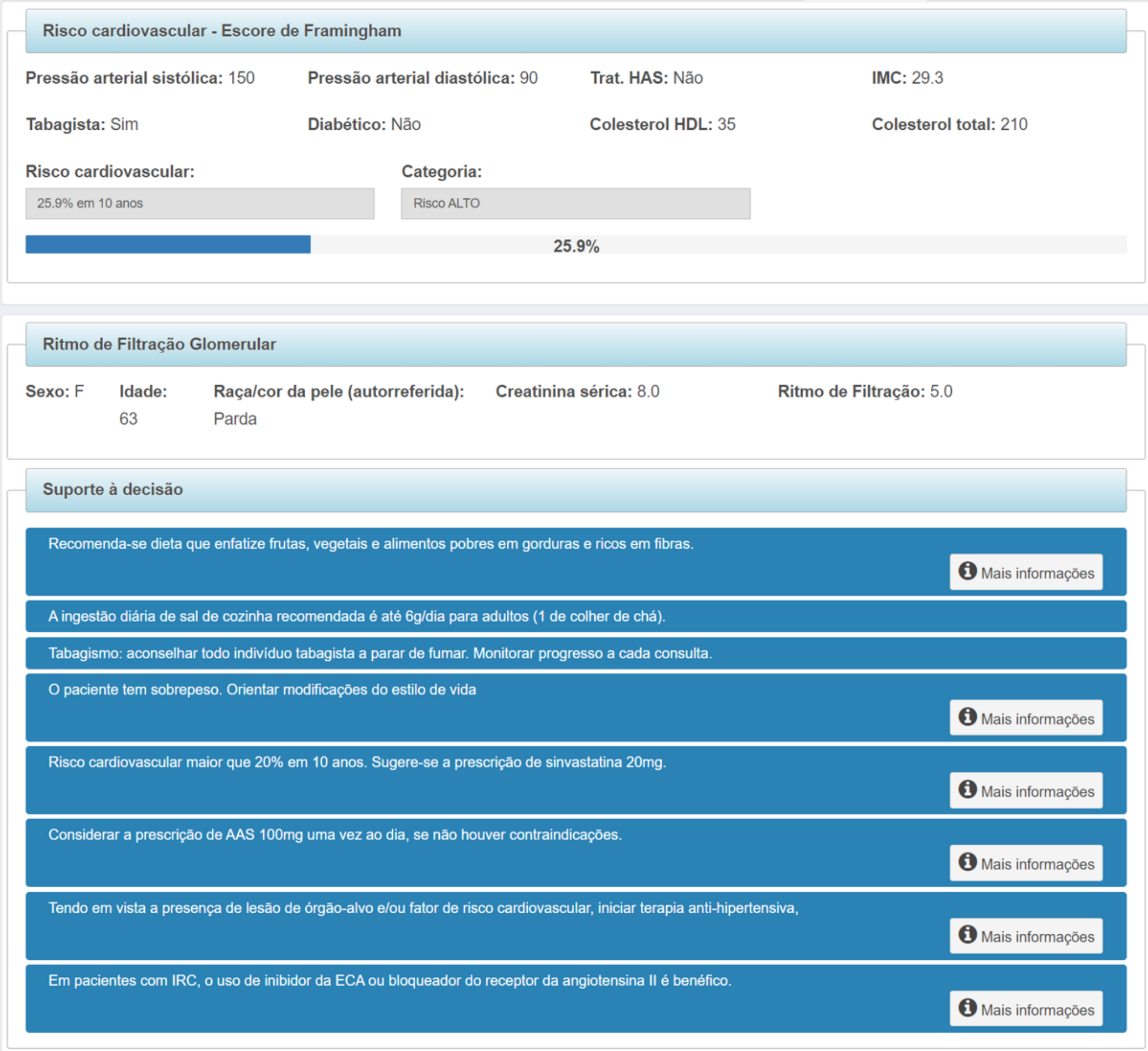
Houve internação? Sim Não preenchido [✓ Não]

Especialidades	Atendido por especialista	Marcar atendido
	Nenhum registro adicionado.	

To develop the clinical decision support functionality, recommendations of Brazilian and international evidence-based guidelines assessing hypertension, diabetes, cardiovascular risk, and chronic renal disease were reviewed [19-29]. The most relevant clinical recommendations for the management of patients with hypertension, diabetes, or both were derived and organized into a decision algorithm, which bases the CDSS alerts, reminders, and suggestions. In case of conflicting recommendations, the one with the best level of evidence was chosen. When evidence-based guideline recommendations were not available or if they were considered outdated, evidence-based summaries were assessed [30]. Previous experiences of our group were important for the development of this functionality. Suggestions received about usability in previous opportunities were decisive to improve and create new messages and resources [17,18].

The clinical decision support tab is shown in Figure 4. In this tab, health care professionals could also access personalized messages to guide the management of each patient, generated according to the data entered in each consultation.

Figure 4. Clinical decision support tab.



Risco cardiovascular - Escore de Framingham

Pressão arterial sistólica: 150 **Pressão arterial diastólica:** 90 **Trat. HAS:** Não **IMC:** 29.3

Tabagista: Sim **Diabético:** Não **Colesterol HDL:** 35 **Colesterol total:** 210

Risco cardiovascular: 25.9% em 10 anos

Categoria: Risco ALTO

25.9%

Ritmo de Filtração Glomerular

Sexo: F **Idade:** 63 **Raça/cor da pele (autorreferida):** Parda **Creatinina sérica:** 8.0 **Ritmo de Filtração:** 5.0

Suporte à decisão

Recomenda-se dieta que enfatize frutas, vegetais e alimentos pobres em gorduras e ricos em fibras.

A ingestão diária de sal de cozinha recomendada é até 6g/dia para adultos (1 de colher de chá).

Tabagismo: aconselhar todo indivíduo tabagista a parar de fumar. Monitorar progresso a cada consulta.

O paciente tem sobrepeso. Orientar modificações do estilo de vida

Risco cardiovascular maior que 20% em 10 anos. Sugere-se a prescrição de sinvastatina 20mg.

Considerar a prescrição de AAS 100mg uma vez ao dia, se não houver contraindicações.

Tendo em vista a presença de lesão de órgão-alvo e/ou fator de risco cardiovascular, iniciar terapia anti-hipertensiva,

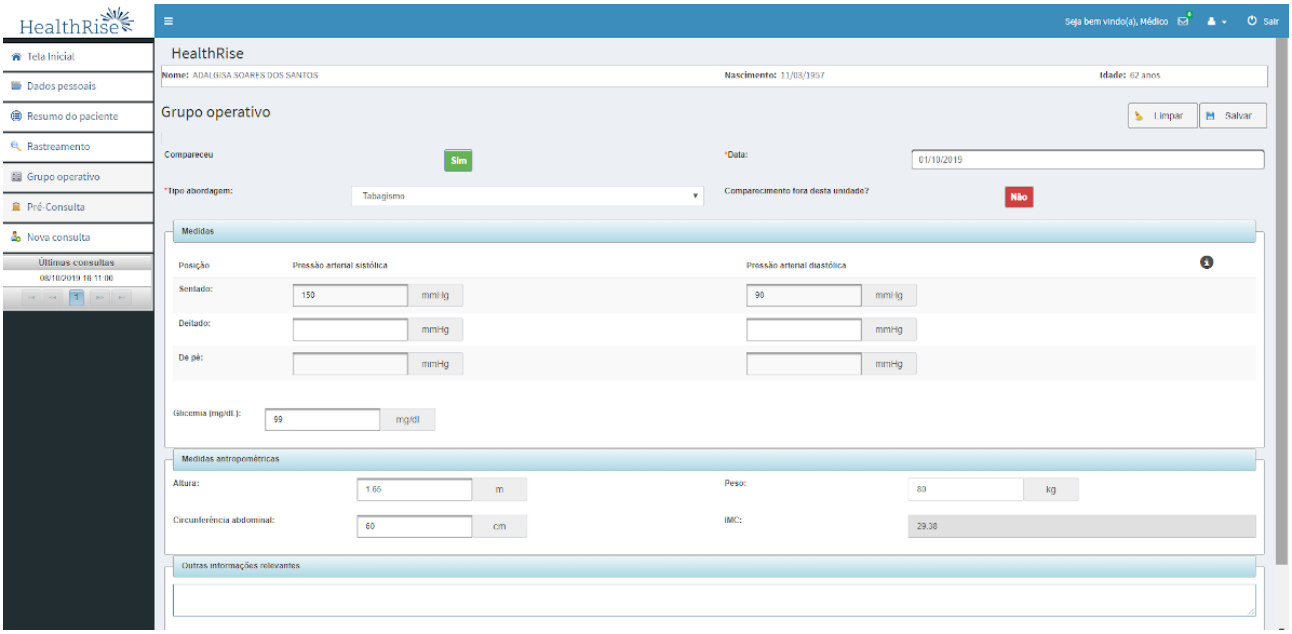
Em pacientes com IRC, o uso de inibidor da ECA ou bloqueador do receptor da angiotensina II é benéfico.

The following variables were calculated using the data entered: (1) BMI, (2) estimated glomerular filtration rate using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [31], and (3) cardiovascular risk based on the Framingham score [32]. For the alert messages, each one contained a summarized recommendation, tailored according to each patient's state of health, with auxiliary text containing detailed information and references that support the recommendation.

In the patient care plan, the health care professional can register nonpharmacological interventions, complementary examinations requested, specialist referral and drug prescription (the last 2 functionalities are available only to doctors), and the date for the next consultation. It is also possible to request teleconsultation, defined as a second opinion system that allows an information exchange between distant and local health care professionals, in order to discuss a clinical case when a specialist is not locally available [33]. For this project, it was possible to forward doubts straight to a family physician and an endocrinologist, who were exclusively available for this project, or to other medical and nonmedical specialties, following the workflow of the Telehealth Network of Minas Gerais, a large public Telehealth service which assists 816 municipalities in the state of Minas Gerais [34]. All teleconsultations were asynchronous. The health care professional could choose which specialty he/she was referring the teleconsultation to, and the normal response time did not exceed 48 hours.

In the educational health groups (Figure 5) screen, the health care professional can register activities performed in different groups in which that patient belongs to (such as hypertension, diabetes, smoking, nutrition).

Figure 5. Health educational groups screen.



HealthRise

Nome: ADALGISA SOARES DOS SANTOS **Nascimento:** 11/03/1957 **Idade:** 62 anos

Grupo operativo

Compareceu: Sim

*Data: 01/10/2019

*Tipo abordagem: Tabagismo

Comparecimento fora desta unidade? Não

Medidas

Posição	Pressão arterial sistólica	Pressão arterial diastólica
Sentado:	150 mmHg	90 mmHg
Deitado:	mmHg	mmHg
De pé:	mmHg	mmHg

Glicemia (mg/dL): 99 mg/dl

Medidas antropométricas

Altura: 1.65 m

Peso: 80 kg

Circunferência abdominal: 60 cm

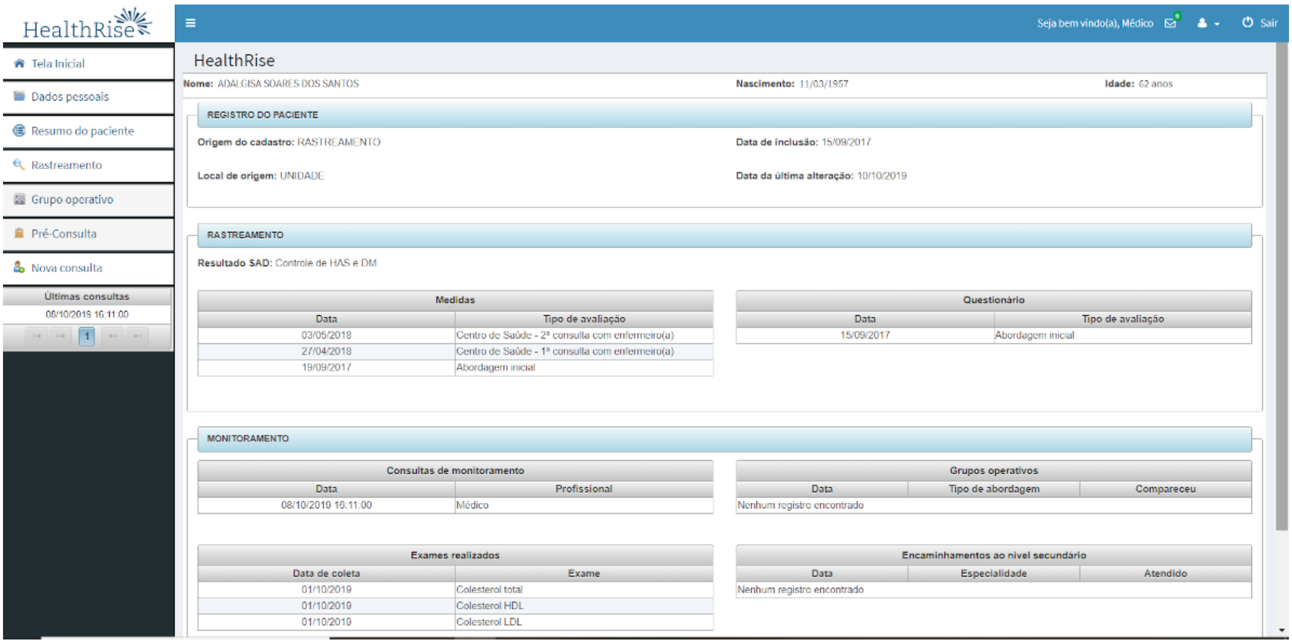
IMC: 29.38

Outras informações relevantes

Patient data obtained in each group (weight, waist circumference, and blood pressure measurements, as well as capillary blood glucose) can be registered, as well as individual absenteeism. There is also a free-text field for any other information that the health professional believes is necessary.

The patient summary screen (Figure 6) shows an overview of the dates of consultations and complementary examinations for each patient.

Figure 6. Patient summary screen.



Últimas consultas: 08/10/2019 16:11:00

HealthRise

Nome: ADALGISA SOARES DOS SANTOS **Nascimento:** 11/03/1957 **Idade:** 62 anos

REGISTRO DO PACIENTE

Origem do cadastro: RASTREAMENTO

Data de inclusão: 15/09/2017

Local de origem: UNIDADE

Data da última alteração: 10/10/2019

RASTREAMENTO

Resultado SAD: Controle de HAS e DM

Medidas

Data	Tipo de avaliação
03/05/2018	Centro de Saúde - 2ª consulta com enfermeiro(a)
27/04/2018	Centro de Saúde - 1ª consulta com enfermeiro(a)
19/09/2017	Abordagem inicial

Questionário

Data	Tipo de avaliação
15/09/2017	Abordagem inicial

MONITORAMENTO

Consultas de monitoramento

Data	Profissional
08/10/2019 16:11:00	Médico

Grupos operativos

Data	Tipo de abordagem	Compareceu
Nenhum registro encontrado		

Exames realizados

Data de coleta	Exame
01/10/2019	Colesterol total
01/10/2019	Colesterol HDL
01/10/2019	Colesterol LDL

Encaminhamentos ao nível secundário

Data	Especialidade	Atendido
Nenhum registro encontrado		

To conclude the consultation, the health care professional must save the data entered and has the option to print a structured medical record file with all information entered in that consultation, to attach to a physical patient chart. The medical prescription and the referral form can also be printed. Data are recorded in the application and transmitted to a telehealth care center as soon as a stable internet connection was available.

In the patient management screen, it is possible to generate patient lists, according to specific monitoring indicators.

Pretesting

The prototype has been tested multiple times with test cases to ensure that the system was operating as intended, free of bugs, and that the recommendation results matched the prespecified decision tree. A manual insertion of data by 2 physicians, 2 pharmacists, 1 nurse, and 3 medical students with test cases was performed to verify the recommendation response suitability.

Expert Panel Assessment

After adjustments, the prototype was submitted for expert panel assessment, composed of 2 cardiologists, 1 primary care physician, 1 endocrinologist, 1 nurse, 1 pharmacist, and 1 physiotherapist, known as technical reference, who were independent from the project and its implementers, in order to retain a degree of impartiality [15]. The experts used the application for 2 weeks, and were asked to assess functional suitability, stability, appropriateness of CDSS content, clinical benefit, readability, strengths, inconsistencies, perception of usefulness, whether it was appropriate to the local context, and satisfaction with the application. The prototype was then re-adjusted with the necessary changes, according to feedback from the expert panel.

CDSS Implementation

The field study was conducted from October 17 to October 18 in 5 primary care centers in Teófilo Otoni, the main city in the northeast region of Minas Gerais, with 141,934 inhabitants [35], and 29 primary centers in 9 small towns (with population less than 20,000 inhabitants), as part of the HealthRise Brazil Project—Vale do Mucuri. The region has a low human development: a mean Human Development Index (HDI) of 0.701 for Teófilo Otoni and 0.540-0.595 for the other municipalities, similar to poor African countries. About one-third of the population lives in rural areas (32.3%) [36]. The HealthRise Brazil Project (2016-2018) consists of 2 parts: patient screening and management (logic model is shown in Multimedia Appendix 2).

The intervention was multifaceted, including extensive training of primary care teams, empowerment of patients through education, improvements in access to examinations, strengthening of health educational groups, and implementation of the computer decision support application.

This study is part of patient management. For this substudy, 34 primary health centers in urban and rural areas were included, with family health teams composed of 1 physician and 1 nurse, 1 nurse assistant, and 4 to 6 full-time CHWs [37]. Family health service (FHS) teams provided comprehensive, universal primary care to defined geographical catchment areas, covering populations of up to 1000 households each (between 3000 and 4000 inhabitants), with no overlap or gap between catchment areas. Each FHS team member had defined roles and responsibilities, and national guidelines helped structure FHS responses to most health problems [37,38]. Pharmacists, physiotherapists, nutritionists, psychologists, physical education specialists, and social workers were part of the multidisciplinary primary care support teams (NASF-AB). Each NASF-AB team assisted 3 or 4 primary care centers [37]. Despite the existence of a university in Teófilo Otoni, clinicians have limited access to specialist referrals and to continuous learning.

Local health authorities and stakeholders were involved in the implementation of the system. The implementation of the application was stepwise, following a schedule of in-person and practical trainings. The participants had the opportunity to handle the devices that would be used. Different training strategies were developed for each professional category. The participants received printed material for further consultation in case of doubt. Furthermore, an online tutorial about the topic was recorded and was available on a platform where they could access online classes about different topics such as hypertension and diabetes management, cardiovascular risk, and operative group planning and organization. Two technology technicians were available online and by telephone to solve doubts and difficulties that arose while using the application. If necessary, they were able to visit each unit.

Fidelity was monitored periodically through checking whether field implementation altered the functionality and stability of the system, changing the intervention from that which was intended, as recommended by the World Health Organization [15]. Malfunctions and other problems with use of the CDSS were quickly fixed.

Feasibility, Usability, and Satisfaction Assessment

A questionnaire evaluating perceived feasibility, usability, and utility of the application and professionals' satisfaction, previously developed by our group [18], was applied after 6 months. The first part of the questionnaire included participant characteristics: age, specialty, time since graduation, time working in primary care, previous knowledge in dealing with health technologies, previous access to the internet, and previous access to continued education. The second part included Likert scale questions, varying from 1 (strongly disagree) to 5 (strongly agree), to assess feasibility, usability, and satisfaction. This instrument was evaluated in its reliability capacity applied in this context.

At the end of the HealthRise Project, 2 focus groups with primary care physicians and nurses were performed. The health care professionals talked about the advantages of using the software and what they thought should be improved. The sessions lasted for 1.5 hours and were audio-recorded. Questions started out broadly and became more specific as participants felt more comfortable and opened up.

Data Analysis

Distributions of continuous variables were examined for normality by using Kolmogorov–Smirnov tests. As the distributions were asymmetrical, these variables were expressed as median and corresponding IQR. Categorical variables were presented as frequencies and proportions. The reliability of the questionnaire to evaluate the perceived feasibility, usability, and utility was evaluated using Cronbach α. Data analyses were performed using the software IBM SPSS Statistics for Windows, version 21.0 (IBM Corp.).

Ethics

Ethical approval was obtained from the Universidade Federal dos Vales do Jequitinhonha e Mucuri Research Ethics Committee (number 65808517.9.0000.5108). Written and informed consent was obtained from all participants.

Results

Phase I: CDSS Development and Validation

An initial bank of 168 reminders and suggestions was created for the clinical decision support functionality. After message content refinement, a final bank of 159 reminders and suggestions was used. Textbox 1 shows examples of the CDSS messages. As there were many messages, we choose to extensively test 26 messages first; we implemented the application with those messages, and continued testing the other messages extensively. We divided the messages in blocks of 10, ranked the groups according to the priority of the recommendations for diabetes and hypertension management, and tested each group before implementation.

Textbox 1. Examples of clinical decision support system reminders and suggestions. ACE: angiotensin-converting enzyme.

- “Reinforce the importance of adherence to treatment and participation in the educational group. Request a visit from the community health worker”.
- “Cardiovascular risk greater than 20% in 10 years. The prescription of statin is suggested.”
- “Do not prescribe thiazide diuretic for this patient due to reduced renal function.”
- “If you decide to add another anti-hypertensive, choose one of the first-line groups (ACE inhibitor, angiotensin receptor II blocker, thiazide diuretic, long-acting calcium channel blocker), taking the contraindications into account.”
- “Metformin is being used at a dosage above the maximum recommendation. Dose reduction is suggested.”
- “This patient did not have a fundoscopic assessment in the last year”.
- “It is suggested to schedule the next consultation in one month, to reassess blood pressure levels”.

During internal tests, most errors identified were due to misplaced parenthesis and connectors in the commands of the decision algorithm. Posterior manual insertion of data evidenced remaining errors, spelling mistakes, and inconsistencies that were properly revised before the test by the panel of specialists. The expert panel–suggested changes were used to enhance the application.

Phase II: Feasibility, Usability, and Satisfaction Assessment

Quantitative Analysis

At the 6-month assessment, there were 1939 patients registered in the application database and 2160 consultations were performed by primary care teams. Of the 96 health care professionals who were invited for the usability assessment, 26% (25/96) were physicians, 46% (44/96) were nurses, and 28% (27/96) were from other health professionals. The number of participants per town varied from 5 (in Crisólita) to 14 (in Ataleia). Participant characteristics are shown in Table 1. Age varied from 23 to 68 and time since graduation varied from 1 to 29 years. Nurses and other health professionals were predominately women.

Table 1. Characteristics of health care professionals who participated in the study (N=96).

Variable	Physicians (n=25)	Nurses (n=44)	Other health professionals (n=27)	Total
Age (years), median (IQR)	35 (30-43)	33 (30-36)	32 (27-36)	33 (30-36)
Female sex, n (%)	11 (44)	33 (75)	22 (81)	66 (69)
Time since graduation (years), median (IQR)	9.0 (7.0-11.0)	9.0 (6.5-11.0)	5.5 (3.0-9.0)	8.0 (4.3-10.8)
Time working in primary care (years), median (IQR)	7.0 (4.0-16.0)	5.5 (2.0-9.4)	3.0 (1.0-4.0)	5.0 (2.0-9.0)
Self-reported knowledge in information technology, n (%)				
Excellent	9 (36)	10 (23)	7 (26)	26 (27)
Good	9 (36)	24 (55)	15 (56)	48 (50)
Satisfactory	7 (28)	9 (20)	5 (19)	21 (22)
Inadequate	0 (0)	1 (2)	0 (0)	1 (1)
Use of any form of technology for work before this research project (yes), n (%)	12 (48)	31 (70)	19 (70)	62 (65)
Computer available in the workplace for routine use before the research project (yes), n (%)	11 (44)	32 (73)	19 (70)	62 (65)
Internet access in the workplace (yes), n (%)	20 (80)	39 (89)	23 (85)	82 (85)
Internet use frequency (yes), n (%)				
Daily	21 (84)	40 (91)	22 (81)	83 (86)
Almost daily	2 (8)	3 (7)	4 (15)	9 (9)
Weekly	2 (8)	0 (0)	0 (0)	2 (2)
Rarely	0 (0)	1 (2)	0 (0)	1 (1)
Participated in an updating activity on management of hypertension, diabetes mellitus, or cardiovascular risk in the last year (yes), n (%)	17 (68)	18 (41)	7 (26)	42 (44)

The results of the questionnaire including 24 items on impressions of feasibility, usability, and utility are presented in Table 2. The reliability analysis of the instrument and 4 dimensions, applied to this population, presented a global Cronbach α of .93 (Table 2 and detailed in Multimedia Appendix 3), showing adequate internal consistency of this instrument for evaluation of the perceived feasibility, usability, and utility of the CDSS.

Table 2. Feasibility, usability, and satisfaction assessment scoresa (N=96).b

Item	Physicians (n=25)	Nurses (n=44)	NASF-AB[c] (n=27)	Total
Feasibility, median (IQR)				
The application can be used in primary care setting	4 (4-5)	5.00 (4-5)	4 (4-5)	4 (4-5)
It is easy to be incorporated in work routine	4 (2-5)	4 (4-4)	4 (3-4)	4 (3-4)
It does not cause significant delays in daily routine	2 (2-3)	3 (2-4)	4 (3-4)	3 (2-4)
Usability, median (IQR)				
Overall, evaluation is good	4 (3.25-4.75)	4 (4-5)	4 (4-5)	4 (4-5)
The screens are easy to understand	4 (4-4)	4 (4-5)	4 (4-4.5)	4 (4-5)
I was able to find the information I was looking for	4 (4-4)	4 (4-4)	4 (4-5)	4 (4-4)
The definitions of comorbidities are clear	4 (4-4)	4 (4-4.75)	4 (4-5)	4 (4-4)
The fields are easy to fill	4 (4-4)	4 (4-5)	4 (4-5)	4 (4-5)
It has proper interface	4 (4-4)	4 (4-5)	4 (4-5)	4 (4-5)
It is stable during the use	4 (2-4)	3 (2-4)	3 (2-4)	3 (2-4)
Utility, median (IQR)				
The application might improve patient care	4 (4-5)	5 (4-5)	5 (4-5)	4 (4-5)
The recommendations brought new information on hypertension, diabetes mellitus, and cardiovascular risk	4 (4-4)	4 (4-5)	4 (4-5)	4 (4-5)
It was useful to calculate the cardiovascular risk	4 (4-4.75)	4 (4-5)	4 (4-5)	4 (4-5)
It was useful to promote cardiovascular disease prevention	4 (4-5)	4 (4-5)	4 (4-5)	4 (4-5)
It assisted me to treat my patients	4 (4-4)	4 (4-5)	4 (4-4)	4 (4-5)
It assisted me to choose complementary examinations	4 (3-4)	4 (4-5)	4 (3.75-4.25)	4 (4-5)
It was helpful to decrease referral to specialists	4 (2-4)	4 (3-4)	4 (4-4)	4 (3-4)
It was useful for different professional categories	4 (3-4)	4 (4-4)	4 (4-4)	4 (4-4)
I believe the recommendations are appropriate	4 (3-4)	4 (4-4)	4 (3.75-4.25)	4 (4-4)
I used the suggestions to modify patient care	4 (4-4)	4 (4-5)	4 (4-5)	4 (4-5)
Satisfaction, median (IQR)				
Overall, I am satisfied with the application	4 (4-4.75)	4 (4-5)	4 (4-5)	4 (3-4)
It may be beneficial for patient care	4 (4-4.75)	4 (4-5)	4 (4-5)	4 (4-5)
I will keep using it for patient care	4 (4-5)	4 (4-5)	4 (4-5)	4 (4-5)
I would recommend this application	4 (3-4.75)	4 (4-5)	4 (4-5)	4 (4-5)

[a] Ranging from 1 to 5.

[b] Cronbach α for the Global questionnaire (24 items) is .93.

[c] NASF-AB: Núcleo Ampliado de Saúde da Família e Atenção Básica (multidisciplinary primary care support teams).

At the end-line assessment, health care professionals registered 4211 patients in the application’s database and 7960 consultations were performed by **primary health care** teams. Patients were predominantly female (2819/4211, 66.94%), median age was 55.0 (IQR 47.0-62.0). A total of 3993 (94.82%) patients were diagnosed with hypertension and 1028 (24.41%) were diagnosed with diabetes and there were 810 (19.24%) patients with diagnosis of both hypertension and diabetes.

Qualitative Analysis

In the focus groups, 13 physicians and 4 nurses participated (71% [12/17] female, median age 34 [IQR 31-39], ranging from 24 to 69 years). With regard to the difficulties dealing with the technology, the majority of participants reported some difficulties in the first days using the software, but no problems afterward. However, there was a statement of more difficulties among older professionals, who received help from the younger team members:

> *I started handling the software for patient assistance, as a nurse. I had no difficulties. Nowadays I help other professionals, such as doctors, who are more used to pen and paper. My unit has a doctor with 30 years experience, who had more difficulty adapting*

> *(using the software). As I showed him step by step and all functionalities of the software, he became interested and found it cool When he learnt that he could write patient prescriptions and print it using the software, he was very satisfied. So now, he prints the patient prescription and the exams, correctly. This way, he saves a lot of time.* [Female, 37-year-old nurse, 13 years' experience working as a primary care nurse]

The structured clinical evaluation was perceived as an advantage, as exemplified below:

> *The questions we always ask (using the software). Are you feeling shortness of breath? Do you wake up out of breath? Sometimes in a rush, we forget (to ask those questions to the patient). But we always ask them (when using the software). It is as if you take the query and say: this is indispensable for you to ask to the patient. You always have to remember to ask about it. So, for me it's great. I will always remember to give counselling, for example, when the patient is a smoker, at the end there is always the 'Did you talk about smoking?' reminder. So there you will remember to tell the patient "Look, you have to stop smoking". For me, this is very cool.* [Female, 31-year-old physician, 4.5 years' experience working as a primary care physician]

> *There was no habit of measuring the blood pressure sitting, lying down and standing. After we started using the software, it became routine. We do not need to say anything. The nurse technician already measures the blood pressure of all patients that way.* [Female, 31-year-old physician, 4.5 years' experience working as a primary care physician]

However, the time spent in the consultation was seen as a challenge for some participants. In some cases, they were forced by the health department to attend each patient for a maximum of 15 minutes, but this time was not sufficient when attending the patients using the software:

> *I take an average of 30 minutes per consultation, i.e. two consultations. This is the same I do with prenatal consultations for pregnant women.* [Male, 39-year-old physician, 6 years' experience working as a primary care physician]

In other cases, there was an organizational issue in the primary care, which made it difficult to attend patients using the software:

> *My primary care center works as an emergency care. So we set aside a day to attend the patients from the project, which is a longer consultation. However, for example, we book ten patients from the program, but 20 show up; six from the project and the rest are patients with acute diseases. It is a challenge to attend the patients using the software, as there are other patients waiting. That makes it difficult.* [Female, 32-year-old physician, 6 years' experience working as a primary care physician]

Often the CHWs were the ones who registered the patients in the software and their educational level was usually low in that region, so some health care professionals interviewed complained that spelling mistakes in patient names made them waste time trying to guess how the CHWs might have registered the patient's name.

With regard to the clinical decision support functionality, all participants were consistent to affirm they were very satisfied with it. They considered the alerts and recommendations very useful for patient care, and reported that it influenced their patient care plan.

> *We usually don't have time to calculate cardiovascular risk and kidney function. The software helps a lot.* [Male, 39-year-old physician, 6 years' experience working as a primary care physician]

> *When attending a patient with high cardiovascular risk, who was taking 40 milligrams of simvastatin, and also amlodipine for hypertension, I received an alert to adjust the dosage of simvastatin to 20 milligrams, as amlodipine may increase the serum concentration of simvastatin and thus increase the risk of adverse effects. I had no idea, I learnt with the alert and changed patient prescription.* [Female, 31-year-old physician, 4.5 years' experience working as a primary care physician]

There was a suggestion to improve the clinical decision support functionality, creating "red flags":

> *I think the abnormal findings or the alerts could be in red. For example, the alert 'This patient is in stage 3a renal failure' should be in red.* [Female, 39-year-old physician, 6 years' experience working as a primary care physician]

Concerning the ability to perform teleconsultations, the majority of participants said they have no time for it during the working hours. The ones who do it usually have to use their off-working hours:

> *I love to do teleconsultations, it helps the patients a lot. So much so that I am currently second (professional who requested the highest number of teleconsultations) in the state of Minas. We can help a lot people in primary care (using the teleconsultations). But I do it (the teleconsultation) at home; That's why it's hard work and a lot of people don't do it, there's no way to do it on the job.* [Male, 69-year-old physician, 29 years' experience working as a primary care physician]

Finally, each software screen was projected and we discussed which variables they felt could be removed. All participants were consistent to affirm that all variables were very useful, except for patient admission, in which they had to fill in the date of admission and patient discharge, as well as the reason for admission. They considered these variables were not necessary, and their work could be reduced if they could be removed.

Discussion

Main Findings

This study described the development, usability, and satisfaction assessment of a web-based CDSS for diabetes, hypertension, and cardiovascular risk management in primary care. Its implementation in a low-resource setting, for users who were used to the paper-based patient records only, has shown to be feasible. Participants with a wide age range and experience in primary care found it to be usable, reported satisfaction, and attended almost 2000 patients using the system over 6 months and over 4000 patients by the end of follow-up.

With the adoption of the 2030 Agenda for Sustainable Development and the 17 Sustainable Development Goals [39], health systems and stakeholders worldwide are interested in innovative approaches to achieving universal health coverage objectives [15,40]. Although the Brazilian Constitution states that health care is free and provided by the public health system to all citizens, patients have unequal access to specialized health care services, especially in remote municipalities. Primary care physicians are often young people (less than 35 years of age) with little professional experience (less than 10 years of medical practice) [41]. Additionally, in those municipalities, there is a strong sense of professional isolation and high turnover of health professionals, which compromises the quality of health care [42]. In a recent meta-analysis, blood pressure control rates ranged from 43.7% to 67.5% in Brazilian patients [43]. Brazil is also one of the “most important examples of the alarming picture of diabetes in emergent societies,” with the fourth largest number of people with diabetes, and a poorer glycemic control rate than that observed in Europe and the United States. Even when a less stringent target was considered, only 48.5% of patients had an HbA1c level less than 8% [44]. In this context, digital health interventions may be very useful, especially those focusing on existing evidence-based health interventions, on the determinant layers of universal health coverage [15].

As CDSS recommendations evolve over time as evidence and the patient’s clinical state evolves, it provides a dynamic and personalized patient care plan that can be easily accessed by any health team member at any point in time [7], and may reduce the likelihood of unhelpful or risky prescriptions [45]. In this study, CDSS was implemented as part of a multifaceted strategy to enhance the delivery of evidence-based care in a resource-constrained area. It is innovative, as the majority of previous studies assessed the use of CDSS for hypertension and diabetes management in high-income countries [46]. Therefore, there is paucity of data on low- and middle-income countries and rural communities, which are the areas in most need and where the access to information technologies may be impaired. In our study, one-third of the total sample (34/96, 35%) and 52% of the doctors (13/25) did not use any form of technology for work before this research project. They did not use any kind of computerized physician order entry or electronic health records, so the implementation of the CDSS brought a significant change in workflow, which could be an important barrier for its usability [9,47]. Therefore, motivating these health care professionals and providing training in using the CDSS were challenging. Developing user-friendly interfaces and avoiding unnecessary information and excess of clicks were essential. Participants agreed that the interface was suitable. Making a tutorial video was important, but it did not replace the need for in-person training, and technical support was of utmost importance. Although the expert panel found the application intuitive, with no need for previous training, the health care professionals did not agree with that, which suggests different levels of technological skills.

With regard to the change in the workflow, health care professionals had to enter all the data items manually. It is reported that the burden of data entry may make them give up and abandon the CDSS [9]. At the same time, there is evidence that arduous data entry facilities adversely affect clinicians’ satisfaction [48], but the specificity level of computer-generated advice is known to highly influence the chance that physicians adhere to the advice [9]. To provide specific advice, is not possible to rely on limited data. A solution we found to try to overcome those barriers, facilitating the integration in the workflow and avoiding duplication of work, was to enable the health care professional to print the consultation using the CDSS and attach it to patient charts. The fact that it could be used by all primary care practitioners, each category with its allowances, and one complementing the other, with the inclusion of the educational groups and a patient management screen, was also important. Additionally, all the data on comorbidities and medications from a previous consultation did not need to be filled in again. Although the majority of primary care practitioners reported satisfaction with the application, doctors gave a low score for the variable “significant delays in daily routine” (median 2.0, IQR 2.0-3.0). Although the median score for doctors for “easy to be incorporated in work routine” was 4.0, 30% (7/23) of them choose a lower score (P25 [first quartile]=2.0). This was also reported in the focus groups. By contrast, when asked about reducing variables in the application, all professionals were consistent with the request to keep it the way it was.

Although the clinical impact assessment was not the focus at the early stages of implementation, there was improvement in outcomes related to hypertension and diabetes management, which are shown in another publication [12]. It is well established that managing the global cardiovascular risk is more important than only reducing blood pressure, glucose levels, or both [49], and this cardiovascular risk should drive preventive strategies. However, the majority of published CDSS did not include the assessment of cardiovascular risk. Choosing the cardiovascular risk calculator to include in the software was challenging, as no score was validated for the South American population [25]. We decided to use the Framingham score [32], as it was widely validated in different populations (European, Asian, Australian, and North American) and assessed important outcomes (coronary artery disease, stroke, peripheral artery disease, and heart failure) with no limitation in number of events (as in the case with the Atherosclerotic Cardiovascular Disease Risk Algorithm [ASCVD], which assessed only the first event) or restricted to fatal events only (as in the case with the Systematic Coronary Risk Evaluation [SCORE]) [50]. Additionally, it includes a number of variables that are easily obtained. In case total cholesterol and high-density lipoprotein cholesterol levels are not available yet, the risk can be estimated using another version of the score, in which BMI replaces the lipid levels [32,51].

Limitations

This is study has some limitations. As was determined by the Ethics Committee, responses to the questionnaire were not identified, so we were not able to assess the results of feasibility, usability, and satisfaction according to the number of patients managed by each user. Additionally, there was no formal evaluation of the effectiveness. The software was part of a multifaceted intervention, whose results were published elsewhere [12], and it is not possible to access the impact of the software by itself. Meta-analyses suggest that CDSS with recommendations to both patients and health care professionals is more effective than providing CDSS to health care professionals only [8,52]. In this study, due to the high levels of illiteracy, we opted not to work directly with patients at that time.

Success Factors

The majority of health care professionals (52/96, 54%) and 32% (8/25) of the doctors had no access to update an activity in the previous year. It is well established that CDSSs are designed to support decision making rather than making the decisions for the user, who makes the final decision based on his/her knowledge, experience, and scientific evidence provided by the CDSS or other sources [47]. Therefore, another important step was clinical training. Professionals had access to online classes, which were designed for each professional category and took into account their reality in terms of resources. Therefore, we believe that the opportunity of updating knowledge via online classes and by using the CDSS may have been one success factor, as has been already reported in the literature [7,8]. For those health care workers who had impaired access to continued education, there was a perception that the CDSS was a useful tool to improve patient care. It may have overcome the delays in daily routine and the perceived difficulty to be incorporated into the clinical workflow.

Management of diabetes and hypertension is a team effort, and many health care professionals may be involved [53,54], so we believe that the fact that all health care practitioners could access the CDSS, each on in their own role, was another important success factor, avoiding keeping patient care completely medical centered.

Careful monitoring of CDSS use rates, requesting feedback from users, and adjusting the application accordingly were other important success factors. Interventions lacking robust monitoring activities are unlikely to generate the impact expected from them and may lead to misjudgment of the digital health intervention as being not effective. As recommended by the WHO, robust monitoring and evaluation plans are essential to support effective implementation and potential intervention scale-up [15].

It is known that the level of adoption of a digital health intervention by users is “dependent on the end-users’ interaction with the technology and their belief/opinion that use of the technology will benefit their health” [16]. The responses to the questionnaire and the comments in the focus groups showed that this was probably another success factor in our study.

To ensure sustainability, the Brazilian Ministry of Health, MG state government, and municipal governments were involved from the beginning, as well as different stakeholders and local health professionals. The successful experience caught the Brazilian Ministry of Health’s interest, who in turn sponsored the continuation of the project’s activities for 2 years in the same region.

Next Steps

We are planning to expand the project to primary care units in other towns, to include other diseases, and to implement in other settings. Our team is now working on the intervention refinement and software improvement, to subsequently conduct a cluster randomized controlled study to test the intervention implementation in other settings. The integration of the CDSS with the incipient electronic medical record of the Brazilian public health system (e-SUS) is a challenge, but we are currently working to overcome this barrier. Although there is limited transferability of the results to other settings, the lessons learnt may be useful for the implementation of other CDSSs in Brazil and around the world.

Conclusion

In this study, a CDSS developed to assist the management of patients with hypertension or diabetes or both was applicable in the context of primary health care setting in a low income region, with good user’s satisfaction and potential to improve adherence to evidence-based practices.

Acknowledgments

HealthRise Brazil Project was designed in partnership with Medtronic Foundation (sponsoring institution), Abt Associates, and Institute for Health Metrics and Evaluation. Those institutions had no influence on software development, implementation, and assessment. AR was supported in part by the Brazilian research agencies CNPQ (*bolsa de produtividade em pesquisa* 310679/2016-8 and IATS 465518/2014-1) and FAPEMIG (*Programa Pesquisador Mineiro* PPM-00428-17 and *Rede de Teleassistência de Minas Gerais* RED-00081-16). AR and MM are members of the National Institute of Science and Technology for Health Technology Assessment (IATS/CNPq). JO received doctoral grant (number 88887.142558/2017-00) from the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) – Finance Code 001, as part of the IATS/CNPq. KA received grants from the IATS (number 372857/2019-1). The authors thank Santino Filoso for revising the manuscript.

Authors' Contributions

AR, CC, VP, and MM were responsible for the research protocol and coordinated the study. MM, JO, JM, LB, and TS participated in the CDSS development and testing. CC, JO, JM, LC, LR, MM, TS and VP participated in the CDSS implementation. CC was responsible for coordination of the research team for the application of the questionnaire. MM performed the data analysis and drafted the manuscript. All other authors reviewed and edited the manuscript. JO organized the manuscript references. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental figure showing the dynamics of the system.
[PNG File , 95 KB - jmir_v23i1e18872_app1.png]

Multimedia Appendix 2
Supplemental figure showing the logic model of the study.
[PNG File , 1775 KB - jmir_v23i1e18872_app2.png]

Multimedia Appendix 3
Supplemental table with detailed data of feasibility, usability and satisfaction assessment scores.
[DOCX File , 22 KB - jmir_v23i1e18872_app3.docx]

References

1. Williams B. High blood pressure in young people and premature death. BMJ 2011 Feb 22;342(feb22 2):d1104-d1104. [doi: 10.1136/bmj.d1104] [Medline: 21343188]
2. Chow CK, Teo KK, Rangarajan S, Islam S, Gupta R, Avezum A, PURE (Prospective Urban Rural Epidemiology) Study investigators. Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. JAMA 2013 Sep 04;310(9):959-968. [doi: 10.1001/jama.2013.184182] [Medline: 24002282]
3. Benjamin E, Blaha M, Chiuve S, Cushman M, Das SR, Deo R, American Heart Association Statistics CommitteeStroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. Circulation 2017 Mar 07;135(10):e146-e603 [FREE Full text] [doi: 10.1161/CIR.0000000000000485] [Medline: 28122885]
4. Blonde L, Aschner P, Bailey C, Ji L, Leiter LA, Matthaei S, Global Partnership for Effective Diabetes Management. Gaps and barriers in the control of blood glucose in people with type 2 diabetes. Diab Vasc Dis Res 2017 May;14(3):172-183 [FREE Full text] [doi: 10.1177/1479164116679775] [Medline: 28467203]
5. Khunti K, Gomes M, Pocock S, Shestakova MV, Pintat S, Fenici P, et al. Therapeutic inertia in the treatment of hyperglycaemia in patients with type 2 diabetes: A systematic review. Diabetes Obes Metab 2018 Feb;20(2):427-437 [FREE Full text] [doi: 10.1111/dom.13088] [Medline: 28834075]
6. Fox J, Patkar V, Chronakis I, Begent R. From practice guidelines to clinical decision support: closing the loop. J R Soc Med 2009 Nov;102(11):464-473 [FREE Full text] [doi: 10.1258/jrsm.2009.090010] [Medline: 19875535]
7. O'Connor P, Sperl-Hillen J, Fazio C, Averbeck B, Rank B, Margolis K. Outpatient diabetes clinical decision support: current status and future directions. Diabet Med 2016 Jun;33(6):734-741 [FREE Full text] [doi: 10.1111/dme.13090] [Medline: 27194173]
8. Roshanov P, Fernandes N, Wilczynski J, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013 Feb 14;346:f657 [FREE Full text] [doi: 10.1136/bmj.f657] [Medline: 23412440]
9. Jaspers M, Smeulers M, Vermeulen H, Peute L. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011 May 01;18(3):327-334 [FREE Full text] [doi: 10.1136/amiajnl-2011-000094] [Medline: 21422100]
10. Wright A, Sittig D, Ash J, Sharma S, Pang J, Middleton B. Clinical decision support capabilities of commercially-available clinical information systems. J Am Med Inform Assoc 2009;16(5):637-644 [FREE Full text] [doi: 10.1197/jamia.M3111] [Medline: 19567796]
11. Kawamoto K, Lobach D. Clinical decision support provided within physician order entry systems: a systematic review of features effective for changing clinician behavior. AMIA Annu Symp Proc 2003:361-365 [FREE Full text] [Medline: 14728195]
12. Medtronic Foundation, Abt Associates, The Institute for Health Metrics and Evaluation. HealthRise Final Report: Expanding Access to Chronic Disease Care through Community Approaches in Four Countries.: Medtronic Foundation; 2019. URL:

http://www.foundation.medtronic.com/content/dam/medtronic_foundation/our_impact/HealthRise%20Final%20Report_FINAL.pdf [accessed 2019-11-22]

13. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
14. Vogel I. Review of the use of 'Theory of Change' in international development. London: DFID; 2012. URL: http://assets.publishing.service.gov.uk/media/57a08a5ded915d3cfd00071a/DFID_ToC_Review_VogelV7.pdf [accessed 2019-11-23]
15. World Health Organization. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. Geneva; 2016. URL: http://apps.who.int/iris/bitstream/handle/10665/252183/9789241511766-eng.pdf [accessed 2019-11-23]
16. World Health Organization. A guide to implementation research in the prevention and control of noncommunicable diseases. Geneva; 2016. URL: http://apps.who.int/iris/bitstream/10665/252626/1/9789241511803-eng.pdf [accessed 2019-11-26]
17. Maia JX, de Sousa LAP, Marcolino MS, Cardoso CS, da Silva JLP, Alkmim MBM, et al. The Impact of a Clinical Decision Support System in Diabetes Primary Care Patients in a Developing Country. Diabetes Technol Ther 2016 Apr;18(4):258-263. [doi: 10.1089/dia.2015.0253] [Medline: 26840128]
18. Silveira D, Marcolino M, Machado E, Ferreira CG, Alkmim MBM, Resende ES, et al. Development and Evaluation of a Mobile Decision Support System for Hypertension Management in the Primary Care Setting in Brazil: Mixed-Methods Field Study on Usability, Feasibility, and Utility. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e9869 [FREE Full text] [doi: 10.2196/mhealth.9869] [Medline: 30907740]
19. American Diabetes Association. Standards of Medical Care in Diabetes - 2017. Diabetes Care 2017;40(Suppl. 1):S1-S135.
20. Whelton P, Carey R, Aronow W, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018 Jun;71(6):1269-1324. [doi: 10.1161/HYP.0000000000000066] [Medline: 29133354]
21. Mancia G, Fagard R, Narkiewicz K, Redón J, Zanchetti A, Böhm M, Task Force Members. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). J Hypertens 2013 Jul;31(7):1281-1357. [doi: 10.1097/01.hjh.0000431740.32696.cc] [Medline: 23817082]
22. James P, Oparil S, Carter B, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). JAMA 2014 Feb 05;311(5):507-520. [doi: 10.1001/jama.2013.284427] [Medline: 24352797]
23. KDIGO. Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney International Supplements 2013;3:1-150 [FREE Full text]
24. Leung AA, Nerenberg K, Daskalopoulou SS, McBrien K, Zarnke KB, Dasgupta K, CHEP Guidelines Task Force. Hypertension Canada's 2016 Canadian Hypertension Education Program Guidelines for Blood Pressure Measurement, Diagnosis, Assessment of Risk, Prevention, and Treatment of Hypertension. Can J Cardiol 2016 May;32(5):569-588. [doi: 10.1016/j.cjca.2016.02.066] [Medline: 27118291]
25. Malachias MVB, Souza WKSB, Plavnik FL, et al. 7th Brazilian Guideline of Arterial Hypertension. Arq Bras Cardiol 2016;107(3 (Suppl.3)):1-83 [FREE Full text]
26. Meireles A, Alves JA, Galil A, et al. Linha-guia de Hipertensão Arterial Sistêmica, Diabetes Mellitus e Doença Renal Crônica. Belo Horizonte: Secetaria de Estado de Saúde de Minas Gerais; 2013. URL: http://www.saude.mg.gov.br/images/documentos/guia_de_hipertensao.pdf [accessed 2018-04-10]
27. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Estratégias para o cuidado da pessoa com doença crônica: diabetes mellitus. Brasília; 2013. URL: http://189.28.128.100/dab/docs/portaldab/publicacoes/caderno_36.pdf [accessed 2018-04-10]
28. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Estratégias para o cuidado da pessoa com doença crônica: hipertensão arterial. Brasília; 2013. URL: http://189.28.128.100/dab/docs/portaldab/publicacoes/caderno_37.pdf [accessed 2018-04-10]
29. Reiner Z, Catapano AL, De Backer G, Graham I, Taskinen MR, Wiklund O, ESC Committee for Practice Guidelines (CPG) 2008-20102010-2012 Committees. ESC/EAS Guidelines for the management of dyslipidaemias: the Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS). Eur Heart J 2011 Jul;32(14):1769-1818. [doi: 10.1093/eurheartj/ehr158] [Medline: 21712404]
30. Dicenso A, Bayley L, Haynes RB. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. Evid Based Nurs 2009 Oct;12(4):99-101. [doi: 10.1136/ebn.12.4.99-b] [Medline: 19779069]
31. Levey A, Stevens L, Schmid C, Zhang YL, Castro AF, Feldman HI, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009 May 05;150(9):604-612 [FREE Full text] [doi: 10.7326/0003-4819-150-9-200905050-00006] [Medline: 19414839]

32. D'Agostino RS, Vasan R, Pencina M, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753. [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
33. Nerlich M, Balas E, Schall T, Stieglitz SP, Filzmaier R, Asbach P, G8 Global Health Applications Subproject 4. Teleconsultation practice guidelines: report from G8 Global Health Applications Subproject 4. Telemed J E Health 2002;8(4):411-418. [doi: 10.1089/15305620260507549] [Medline: 12626110]
34. Soriano Marcolino M, Minelli Figueira R, Pereira Afonso Dos Santos J, Silva Cardoso C, Luiz Ribeiro A, Alkmim MB. The Experience of a Sustainable Large Scale Brazilian Telehealth Network. Telemed J E Health 2016 Nov;22(11):899-908. [doi: 10.1089/tmj.2015.0234] [Medline: 27167901]
35. Instituto Brasileiro de Geografia e Estatística. Estatísticas sociais por cidade e estado. Teófilo Otoni, Minas Gerais.: IBGE; 2018. URL: http://cidades.ibge.gov.br/brasil/mg/teofilo-otoni/panorama [accessed 2018-04-10]
36. Instituto Brasileiro de Geografia e Estatística. População residente, por situação do domicílio, sexo e idade, segundo a condição no domicílio e compartilhamento da responsabilidade pelo domicílio.: IBGE; 2019. URL: http://sidra.ibge.gov.br/Tabela/1378#resultado [accessed 2019-09-18]
37. Macinko J, Harris M. Brazil's family health strategy--delivering community-based primary care in a universal health system. N Engl J Med 2015 Jun 04;372(23):2177-2181. [doi: 10.1056/NEJMp1501140] [Medline: 26039598]
38. Macinko J, Harris M, Rocha M. Brazil's National Program for Improving Primary Care Access and Quality (PMAQ): Fulfilling the Potential of the World's Largest Payment for Performance System in Primary Care. J Ambul Care Manage 2017;40 Suppl 2 Supplement, The Brazilian National Program for Improving Primary Care Access and Quality (PMAQ):S4-S11 [FREE Full text] [doi: 10.1097/JAC.0000000000000189] [Medline: 28252498]
39. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development. New York: United Nations; 2015. URL: http://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf [accessed 2019-11-25]
40. Kieny M, Bekedam H, Dovlo D, Fitzgerald J, Habicht J, Harrison G, et al. Strengthening health systems for universal health coverage and sustainable development. Bull World Health Organ 2017 Jul 01;95(7):537-539 [FREE Full text] [doi: 10.2471/BLT.16.187476] [Medline: 28670019]
41. Miotto B, Guilloux A, Cassenote A, Mainardi G, Russo G, Scheffer M. Physician's sociodemographic profile and distribution across public and private health care: an insight into physicians' dual practice in Brazil. BMC Health Serv Res 2018 Apr 23;18(1):299 [FREE Full text] [doi: 10.1186/s12913-018-3076-z] [Medline: 29688856]
42. Paim J, Travassos C, Almeida C, Bahia L, Macinko J. The Brazilian health system: history, advances, and challenges. Lancet 2011 May 21;377(9779):1778-1797. [doi: 10.1016/S0140-6736(11)60054-8] [Medline: 21561655]
43. Picon R, Dias-da-Costa J, Fuchs F, Olinto M, Choudhry N, Fuchs S. Hypertension Management in Brazil: Usual Practice in Primary Care-A Meta-Analysis. Int J Hypertens 2017;2017:1274168 [FREE Full text] [doi: 10.1155/2017/1274168] [Medline: 28751987]
44. de Almeida-Pititto B, Dias M, de Moraes ACF, Ferreira S, Franco D, Eliaschewitz F. Type 2 diabetes in Brazil: epidemiology and management. Diabetes Metab Syndr Obes 2015;8:17-28 [FREE Full text] [doi: 10.2147/DMSO.S72542] [Medline: 25609989]
45. Koopman RJ, Kochendorfer KM, Moore JL, Mehr DR, Wakefield DS, Yadamsuren B, et al. A diabetes dashboard and physician efficiency and accuracy in accessing data needed for high-quality diabetes care. Ann Fam Med 2011;9(5):398-405 [FREE Full text] [doi: 10.1370/afm.1286] [Medline: 21911758]
46. Jia P, Zhao P, Chen J, Zhang M. Evaluation of clinical decision support systems for diabetes care: An overview of current evidence. J Eval Clin Pract 2019 Feb;25(1):66-77. [doi: 10.1111/jep.12968] [Medline: 29947136]
47. Castillo R, Kelemen A. Considerations for a successful clinical decision support system. Comput Inform Nurs 2013 Jul;31(7):319-26; quiz 327. [doi: 10.1097/NXN.0b013e3182997a9c] [Medline: 23774450]
48. Shiffman R, Liaw Y, Brandt C, Corb G. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. J Am Med Inform Assoc 1999;6(2):104-114 [FREE Full text] [doi: 10.1136/jamia.1999.0060104] [Medline: 10094063]
49. Stewart J, Manmathan G, Wilkinson P. Primary prevention of cardiovascular disease: A review of contemporary guidance and literature. JRSM Cardiovasc Dis 2017;6:2048004016687211 [FREE Full text] [doi: 10.1177/2048004016687211] [Medline: 28286646]
50. Damen J, Pajouheshnia R, Heus P, Moons KGM, Reitsma JB, Scholten RJPM, et al. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. BMC Med 2019 Jun 13;17(1):109 [FREE Full text] [doi: 10.1186/s12916-019-1340-7] [Medline: 31189462]
51. Gaziano T, Young C, Fitzmaurice G, Atwood S, Gaziano J. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. Lancet 2008 Mar 15;371(9616):923-931 [FREE Full text] [doi: 10.1016/S0140-6736(08)60418-3] [Medline: 18342687]
52. Collins S, Drew P, Watt I, Entwistle V. 'Unilateral' and 'bilateral' practitioner approaches in decision-making about treatment. Soc Sci Med 2005 Dec;61(12):2611-2627. [doi: 10.1016/j.socscimed.2005.04.047] [Medline: 16006027]

53. McGill M, Felton A, Global Partnership for Effective Diabetes Management. New global recommendations: a multidisciplinary approach to improving outcomes in diabetes. Prim Care Diabetes 2007 Feb;1(1):49-55. [doi: 10.1016/j.pcd.2006.07.004] [Medline: 18632019]
54. Sousa W, Vitorino P, Sousa A, Jardim T, Souza W, Moreira H, et al. PP.30.35. Journal of Hypertension 2015;33:e408 [FREE Full text] [doi: 10.1097/01.hjh.0000468659.75999.a7]

Abbreviations

CDSS: clinical decision support system
CHW: community health worker
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration
FHS: family health service
HDI: Human Development Index
NASF-AB: Núcleo Ampliado de Saúde da Família e Atenção Básica

---

Edited by G Eysenbach; submitted 24.03.20; peer-reviewed by S Omboni, S Sarbadhikari; comments to author 25.07.20; revised version received 19.09.20; accepted 21.09.20; published 11.01.21.

Please cite as:
Marcolino MS, Oliveira JAQ, Cimini CCR, Maia JX, Pinto VSOA, Sá TQV, Amancio K, Coelho L, Ribeiro LB, Cardoso CS, Ribeiro AL
Development and Implementation of a Decision Support System to Improve Control of Hypertension and Diabetes in a Resource-Constrained Area in Brazil: Mixed Methods Study
J Med Internet Res 2021;23(1):e18872
URL: http://www.jmir.org/2021/1/e18872/
doi:10.2196/18872
PMID:33427686

Original Paper

Use of BERT (Bidirectional Encoder Representations from Transformers)-Based Deep Learning Method for Extracting Evidences in Chinese Radiology Reports: Development of a Computer-Aided Liver Cancer Diagnosis Framework

Honglei Liu[1,2], PhD; Zhiqiang Zhang[1,2], BS; Yan Xu[3], MD; Ni Wang[1,2], BS; Yanqun Huang[1,2], BS; Zhenghan Yang[3], MD; Rui Jiang[4], PhD; Hui Chen[1,2], PhD

[1]School of Biomedical Engineering, Capital Medical University, Beijing, China

[2]Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China

[3]Department of Radiology, Beijing Friendship Hospital, Capital Medical University, Beijing, China

[4]Ministry of Education Key Laboratory of Bioinformatics, Research Department of Bioinformatics at the Beijing National Research Center for Information Science and Technology, Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing, China

Corresponding Author:
Hui Chen, PhD
School of Biomedical Engineering, Capital Medical University
No 10, Xitoutiao, Youanmen, Fengtai District
Beijing
China
Phone: 86 010 83911545
Email: chenhui@ccmu.edu.cn

Abstract

Background: Liver cancer is a substantial disease burden in China. As one of the primary diagnostic tools for detecting liver cancer, dynamic contrast-enhanced computed tomography provides detailed evidences for diagnosis that are recorded in free-text radiology reports.

Objective: The aim of our study was to apply a deep learning model and rule-based natural language processing (NLP) method to identify evidences for liver cancer diagnosis automatically.

Methods: We proposed a pretrained, fine-tuned BERT (Bidirectional Encoder Representations from Transformers)-based BiLSTM-CRF (Bidirectional Long Short-Term Memory-Conditional Random Field) model to recognize the phrases of APHE (hyperintense enhancement in the arterial phase) and PDPH (hypointense in the portal and delayed phases). To identify more essential diagnostic evidences, we used the traditional rule-based NLP methods for the extraction of radiological features. APHE, PDPH, and other extracted radiological features were used to design a computer-aided liver cancer diagnosis framework by random forest.

Results: The BERT-BiLSTM-CRF predicted the phrases of APHE and PDPH with an F1 score of 98.40% and 90.67%, respectively. The prediction model using combined features had a higher performance (F1 score, 88.55%) than those using APHE and PDPH (84.88%) or other extracted radiological features (83.52%). APHE and PDPH were the top 2 essential features for liver cancer diagnosis.

Conclusions: This work was a comprehensive NLP study, wherein we identified evidences for the diagnosis of liver cancer from Chinese radiology reports, considering both clinical knowledge and radiology findings. The BERT-based deep learning method for the extraction of diagnostic evidence achieved state-of-the-art performance. The high performance proves the feasibility of the BERT-BiLSTM-CRF model in information extraction from Chinese radiology reports. The findings of our study suggest that the deep learning–based method for automatically identifying evidences for diagnosis can be extended to other types of Chinese clinical texts.

(J Med Internet Res 2021;23(1):e19689) doi:10.2196/19689

KEYWORDS

BiLSTM-CRF; natural language processing; radiology reports; information extraction; computer-aided diagnosis; BERT

Introduction

In the past decades, electronic health records (EHRs) from millions of patients have become massive sources of valuable clinical data. Machine learning–based algorithms, especially deep learning algorithms, have been applied effectively to analyze patient data and they have shown promising results, thereby advancing medical research and better informing clinical decision making for the secondary use of EHRs [1,2]. Owing to the high dimensionality, noise, heterogeneity, random errors, and systematic biases, the data mining of EHRs remains challenging. Natural language processing (NLP) technologies could extract meaningful information, thus facilitating the application of clinical texts. There are 2 types of methods for information extraction, namely, rule-based methods and machine learning methods [1]. The use of machine learning methods for data mining of EHRs can derive previously unknown clinical insights and be applied powerfully in clinical decision-making and computer-aided diagnosis of diseases [3,4]. Recently, deep learning methods have had a profound impact in various areas of research because of their simplicity, efficient processing, and state-of-the-art results [5,6]. In particular, recurrent neural networks and Word2Vec embedding are the most popular methods that are utilized in clinical NLP tasks [2]. Deep learning methods have made improvements in various clinical applications, especially for text classification, named-entity recognition (NER), relation extraction, and question answering [7,8]. With growing acceptance and increasing number of applications, deep learning methods have become a baseline in many clinical NLP tasks.

Word embedding is an essential step for sequencing labelling tasks. Learning word representations from massive unannotated documents is a long-established method. The Word2Vec method [9] is the first word embedding approach based on deep learning methods. The model derives the semantic and synthetic meaning of a word on account of its adjacent words by using unsupervised learning. Global Vector word representation [10] is another effective word embedding method, which constructs a global word-word co-occurrence matrix and utilizes matrix factorization to learn embeddings in a lower dimensional space. However, the word-level representations have a limitation that only a single embedding is provided for a word with no thought for polysemy under different contexts. Unlike the traditional embedding methods, ELMo (Embeddings from Language Models) [11] uses a bidirectional language model to embed the context information into word representations. BERT (Bidirectional Encoder Representations from Transformers) [12] is another prominent contextualized word representation model, which uses a masked language model that predicts randomly masked words in a context sequence. Different from ELMo, BERT targets different training objectives and uses a masked language model to learn bidirectional representations. For clinical sequence labelling tasks such as NER, rule-based approach and conditional random fields (CRFs) have been used widely. Deep learning technologies substantially improve the NER performance through multi-layer data representations. Of the popular deep learning methods, BiLSTM (bidirectional long short-term memory) can capture long-range related information effectively. Furthermore, BiLSTM with CRF, known as BiLSTM-CRF, outperforms the traditional models with feature extraction and reduces the workload of feature selection [13].

Due to the difference in the grammatical features from English and the limitation of the EHR corpus, information extraction of Chinese EHRs using NLP remains challenging. In the medical field, researchers have developed information extraction algorithms for varied implementations, including diagnostic models for different diseases such as cancers [14] and childhood diseases [15]. For Chinese NER tasks, BiLSTM-CRF is the most common and practical approach [16,17]. BERT has also received extensive attention in Chinese EHRs. Zhang et al used fine-tuning BERT for NER and relation extraction in several types of Chinese clinical documents. The comprehensive clinical information related to breast cancer was extracted [14]. Wu et al developed an aided clinical diagnosis service on EHRs by using a deep learning model [3]. Liang et al applied an automatic NLP system and achieved a high diagnostic accuracy in childhood diseases [15].

The radiology report is a crucial component of EHRs, as it is the communication bridge between radiologists and physicians. The accuracy and efficiency of diagnosis are limited since it is formulated based on subjective judgment, especially for inexperienced physicians. Hence, extracting useful radiological information from radiology reports has considerable significance in advancing radiological research and clinical practice [18,19]. NLP technologies have received great attention in the processing of radiology reports and have been successfully applied in identifying biomedical concepts [20], extracting recommendations [21], determining the change level of clinical findings [22], and so on.

With the development of machine learning methods in recent eras, computer-aided early diagnosis for cancer based on massive clinical data becomes feasible. Many diseases have been investigated to date, such as hepatocellular cancer [23] and colorectal cancer [24]. In this study, we focused on the computer-aided diagnosis of liver cancer, which remains to be a substantial disease burden in China. For liver cancer diagnosis, dynamic contrast-enhanced computed tomography (CT) is one of the primary diagnostic tests. Imaging findings of the key enhanced scan phases such as the arterial phase, portal phase, and delayed phase are recorded in the radiology reports. According to the guidelines of the Chinese Society of Clinical Oncology (CSCO), hyperintense enhancement in the arterial phase (APHE) and hypointense enhancement in the portal and delayed phases (PDPH) are significant diagnostic evidences for liver cancer [25].

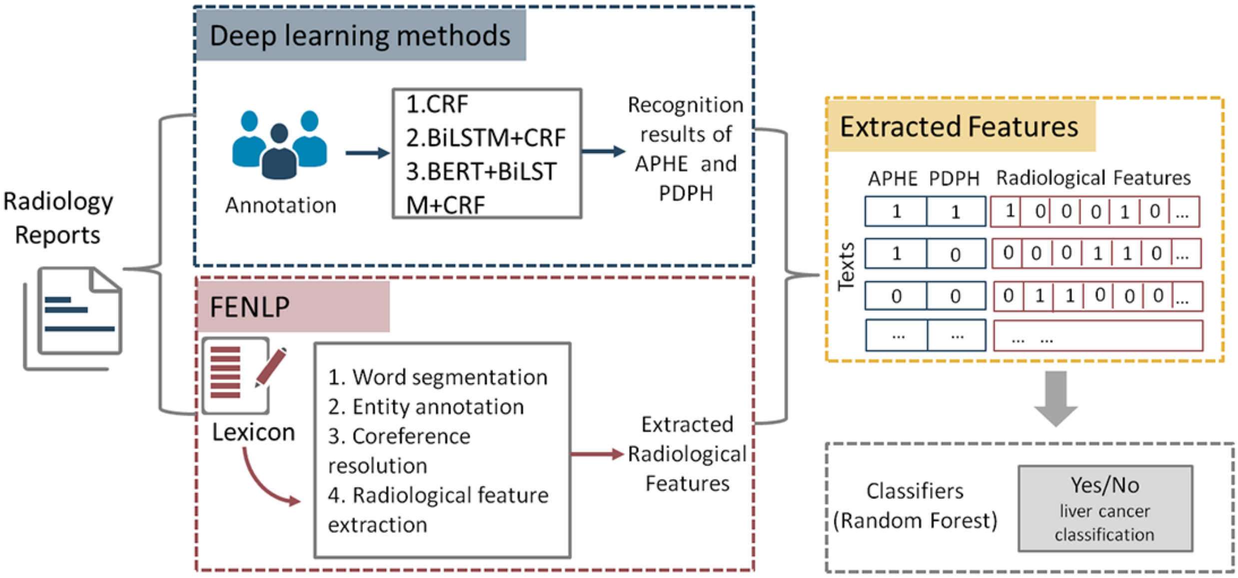
In this study, we designed deep learning–based methods to identify evidences for liver cancer diagnosis automatically. We recognized the phrases of APHE and PDPH by using a BERT-BiLSTM-CRF model by combining a pretrained, fine-tuned BERT language model with BiLSTM-CRF. We also applied the FENLP (feature extraction using the rule-based NLP) method based on the content of radiology reports to extract the radiological features. Therefore, the evidences for diagnosis, considering both clinical knowledge and radiology findings, contained APHE, PDPH, and radiological features extracted by FENLP [26]. With these evidences, we designed a computer-aided liver cancer diagnosis framework by using random forest.

Methods

Evidence Extraction for Diagnosis of Liver Cancer

Figure 1 shows the workflow of the evidence extraction for the diagnosis of liver cancer. We implemented 2 feature extraction methods based on clinical knowledge and the content of radiology reports to generate a radiological feature set. Then, we built a random forest model to predict liver cancer by using these features as inputs.

Figure 1. The workflow of this research. Labels 0/1 represent the absence/presence of a certain feature. BERT: Bidirectional Encoder Representations from Transformers; BiLSTM: bidirectional long short-term memory; CRF: conditional random field; APHE: hyperintense enhancement in the arterial phase; PDPH: hypointense in the portal and delayed phases.



Data Sets

We collected abdomen and pelvic CT radiology reports from a tertiary hospital in Beijing, China, between 2012 and 2019. To protect patient privacy, we removed all the identifying information. An unstructured radiology report has different sections, including Type of Examination, Clinical History, Comparison, Technique, Imaging Findings, and Impressions. The Impressions section summarizes crucial radiology findings from the Findings section and contains a diagnosis indicated by a radiologist. In this study, the diagnosis of liver cancer was determined according to the Impression section and the annotation of experienced radiologists, resulting in 480 patients with liver cancer. We randomly selected 609 patients without liver cancer from our data set. Therefore, 480 and 609 radiology reports for patients with and without liver cancer, respectively, were used in this study. We then trained and evaluated an NER model on the Imaging Findings section. The reports were randomly divided into the training set and the test set in a ratio of 8:2.

BERT-BiLSTM-CRF for Recognition of APHE and PDPH

We considered the recognition of APHE and PDPH as a sequence labelling task at the character level, where the goal was to assign the BIO (Begin, Inside, Outside) tags to each Chinese character. In this study, BIO tags contained B-APHE, I-APHE, B-PDPH, I-PDPH, and O-Outside. We invited 2 radiologists with more than 5 years of medical experience to annotate all the data. If there was any inconsistency, another experienced radiological expert was then asked to make the final annotation, to obtain the gold standard annotated data. To ensure the consistency of the annotation, radiologists were trained in advance. At the report level, APHE and PDPH were not mutually exclusive, that is, 1 report could contain both APHE and PDPH. Of all the reports, 602 had the phrase of APHE and 330 had the phrase of PDPH. For the 480 reports diagnosed with liver cancer, the numbers of APHE and PDPH were 442 and 330, respectively.

BiLSTM-CRF is commonly used in the sequence labeling task. To further improve the recognition performance for the features of APHE and PDPH, we performed the BERT-BiLSTM-CRF model comprising a fine-tuned BERT language model for word embedding and BiLSTM-CRF method for feature recognition. CRF and BiLSTM-CRF model were applied as the baseline. APHE and PDPH in Chinese radiology reports had a variety of presentations such as detailed presentation, CT values of different phases, and abbreviations (Table 1). The deep learning model for the recognition of APHE and PDPH consisted of 3 layers, namely, the word embedding layer, BiLSTM layer, and CRF layer (Figure 2).

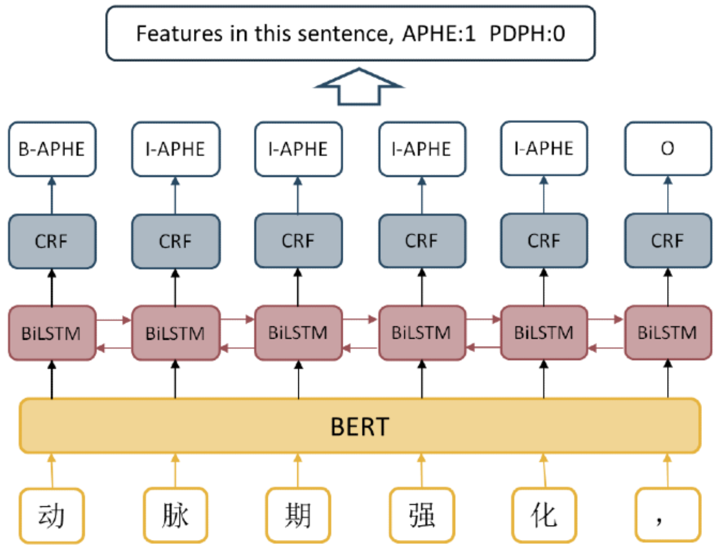
Table 1. Some expressions of APHE[a] and PDPH[b] in Chinese.

Expressions of APHE and PDPH in Chinese	Detailed descriptions
增强后动脉期明显不均匀强化	The arterial phase shows the heterogeneous density in the enhanced scan.
动脉期强化明显	Marked enhancement is shown in the arterial phase.
动脉期可见多发强化灶	Multiple enhancement areas are seen in the arterial phase.
门脉期相对低密度	The portal phase has relatively low density.
门脉期可见消退	PDPH occurs in the portal phase.

[a] APHE: hyperintense enhancement in the arterial phase.

[b] PDPH: hypointense in the portal and delayed phases.

Figure 2. The architecture of the BERT-BiLSTM-CRF model for the recognition of APHE and PDPH. BERT: Bidirectional Encoder Representations from Transformers; BiLSTM: bidirectional long short-term memory; CRF: conditional random field; APHE: hyperintense enhancement in the arterial phase; PDPH: hypointense in the portal and delayed phases.



Word Embedding Layer

The word embedding layer could map and transform the discrete characters into distributed representations. A word-level vector representation learned a real valued vector to represent a word from a large amount of unannotated text. On most NLP tasks, BERT could achieve state-of-the-art performance while requiring minimal architectural modification [27]. In this study, we applied Word2Vec and BERT to train the character vectors, followed by a comparison of the results. The Word2Vec was used with a dimension size of 100 and a batch size of 120. The Word2Vec was pretrained on the Chinese Wikipedia data. The sentence embedding had been pretrained and fine-tuned by BERT on the original Google BERT GitHub repository [28]. The maximum sequence length was set to 256 with a batch size of 64.

BiLSTM Layer

Recurrent neural networks is a family of neural networks, which is usually used for modelling sequential data. The LSTM (Long Short-Term Memory Networks) is a variant of the recurrent neural networks, and it can effectively capture high dependencies and retrieve rich global information. LSTM solves the problem by using the gating mechanism. An LSTM unit consists of 3 gates (ie, Input Gate, Output Gate, and Forget Gate), which can select semantic information in a neural network. Compared with LSTM, BiLSTM can learn forward and backward information of input words by splitting the neurons into 2 directions of a text sequence. We set the number of hidden units in BiLSTM to 100 and the optimizer to Adam.

CRF Layer

For the sequence labelling step in our study, adjacent tags had dependencies. For example, an inside tag I must follow a begin tag B. We applied the sequential CRF to calculate optimal sequence combinations on top of the BiLSTM layer that could consider the dependencies of adjacent tags.

APHE and PDPH Labels at the Report Level

Considering the characteristics of Chinese language and also avoiding the noise, we defined the following as APHE or PDPH features at the report level: (1) 2 continuous characters that were the abbreviations of APHE (ie, 快进) or PDPH (ie, 快出); (2) more than 3 continuous characters that were predicted as APHE or PDPH. Criterion (1) was checked first and was only based on the characters. If not met, criterion (2) was checked, which was based on CRF results.

FENLP for Radiological Feature Extraction

We implemented the NLP pipeline in the Findings section to extract useful features from the unstructured radiology reports to facilitate liver cancer diagnosis. As shown in Figure 1, the

NLP pipeline consisted of 4 successive steps, that is, word segmentation, entity annotation, coreference resolution, and relationship extraction, resulting in radiological features consisting of 1 or more terms. The detailed description of the pipeline is provided in our previous study [26]. The whole pipeline was based on a lexicon that was constructed manually according to Chinese grammatical characteristics. A small number of reports were sampled randomly for generating the lexicon by manual reading. The lexicon contained clinical terms and lists of synonyms. The lexicon was collected in the same hospital and clinical text type with this study. Five entity types (Location, Morphology, Density, Enhancement, and Modifier) were recognized. After coreference resolution, according to the synonym list in the lexicon, we then used several patterns of entity types as rules to obtain the final radiological features (Table S1 of Multimedia Appendix 1). Therefore, the radiological features could be seen as a combination of several entities such as 肝脏+低密度影 (liver + low density) and 肝脏+增强扫描未见强化 (liver + enhancement scan showed no enhancement).

Prediction Models

Using the radiological features obtained by BERT-BILSTM-CRF and FENLP, we used a random forest model for the liver cancer diagnosis. Random forest is an ensemble learning method constructed with a multitude of decision trees, which is widely used in classification tasks. The performance was measured by the recall, precision, and F1 score. Random forest could generate the feature importance score, which was computed by Gini impurity. Gini impurity is a measurement of the probability that a sample is classified incorrectly without a specific feature. In our study, the higher the feature importance score of the radiological features was, the more linked it was with the liver cancer diagnosis. We used the feature importance score to rank all the radiological features.

Results

We extracted the features of APHE and PDPH by using 3 different models, that is, CRF, BiLSTM-CRF, and BERT-BiLSTM-CRF. The recognition results were presented both at the report level and character level (Table 2). At the report level, the performance was computed depending on whether the radiology reports contained a feature of APHE or PDPH. At the character level, the recognition results of BIO tags for each Chinese character were counted. For the character-level recognition results of APHE and PDPH, the BERT-BiLSTM-CRF model obtained the best performance, with F1 scores of 89.14% and 82.19%, respectively. At the report level, the BERT-BiLSTM-CRF model also achieved the best performance (F1 scores of 98.40% for APHE and 90.67% for PDPH). For the other 2 baseline models, the BiLSTM-CRF model outperformed the CRF model but underperformed the BERT-BiLSTM-CRF model. If a single character was recognized as a feature, it would be regarded as noisy information, thereby leading to its exclusion from the report-level results. As a result, the recognition performances at the report level were higher than those at the character level in all the models. We chose the recognition results of APHE and PDPH at the report level by the BERT-BiLSTM-CRF model as the predictors for further liver cancer diagnosis.

The feature extraction method FENLP used the lexicon described in our previous study, which included 831 words and 5 entity types. Entity combinations conforming to specific entity patterns were formulated as radiological features. The patterns included Location + Density, Location + Enhancement, Location + Enhancement + Modifier, Location + Density + Modifier, and Location + Morphology. We retained the radiological features that occurred more than twice. We finally obtained 301 radiological features; among them, 6 features had a frequency higher than 300 (Table S2 of Multimedia Appendix 1).

Table 2. Recognition performance of APHE[a] and PDPH[b] by using different models at the character level and report level.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Character level				
Conditional random field				
APHE	96.05	84.13	72.19	77.70
PDPH	97.44	80.37	59.02	68.06
Bidirectional long short-term memory-conditional random field				
APHE	97.54	90.86	82.56	86.51
PDPH	98.24	84.56	75.72	79.89
BERT[c]+ Bidirectional long short-term memory-conditional random field				
APHE	97.97	91.14	87.22	89.14
PDPH	98.46	88.60	76.64	82.19
Report level				
Conditional random field				
APHE	94.52	98.28	91.94	95.00
PDPH	89.00	87.69	79.17	83.21
Bidirectional long short-term memory-conditional random field				
APHE	95.89	97.30	94.74	96.00
PDPH	93.61	92.19	86.76	89.39
BERT+ Bidirectional long short-term memory-conditional random field				
APHE	98.17	97.62	99.19	98.40
PDPH	93.61	87.18	94.44	90.67

[a] APHE: hyperintense enhancement in the arterial phase.

[b] PDPH: hypointense in the portal and delayed phases.

[c] BERT: Bidirectional Encoder Representations from Transformers.

According to the presence or absence of each feature extracted from either BERT-BILSTM-CRF or FENLP, each radiology report was represented by a 0-1 vector. The prediction results using different patterns of features are shown in Table 3. F1 scores of random forest using features from BERT-BILSTM-CRF and FENLP were 84.88% and 83.92%, respectively. With a combination of both kinds of features, the final F1 score of prediction model increased to 88.55%. Among all the feature input patterns, the precision and accuracy also obtained the highest value while inputting all the features, while the prediction model had the highest recall rate with only 2 features of APHE and PDPH. Among the features with a frequency higher than 10 in all the reports, the top 10 features linked with the liver cancer diagnosis were identified with the feature importance score computed by Gini impurity (Figure 3). The top 2 features were APHE and PDPH, which had substantially larger feature importance scores than the other features extracted from FENLP.

Table 3. Performance of different patterns of features for liver cancer diagnosis.

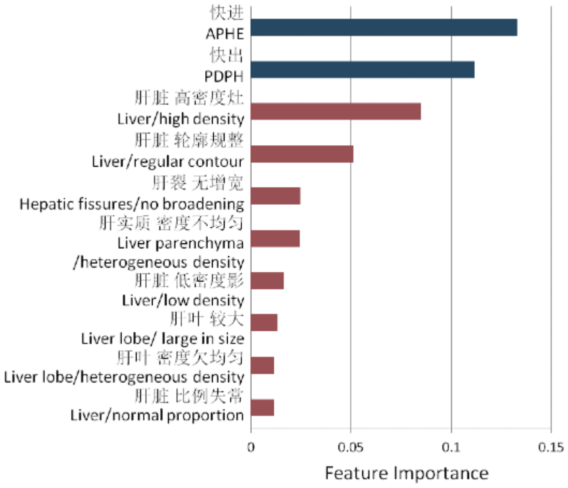
Patterns of Features	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
APHE[a] and PDPH[b]	86.11	81.38	88.70	84.88
Radiological features from FENLP[c]	85.71	83.06	84.80	83.92
All features	90.25	91.52	85.77	88.55

[a] APHE: hyperintense enhancement in the arterial phase.

[b] PDPH: hypointense in the portal and delayed phases.

[c] FENLP: feature extraction using the rule-based natural language processing.

Figure 3. Top 10 radiological features linked with liver cancer diagnosis ranked by feature importance score. APHE: hyperintense enhancement in the arterial phase; PDPH: hypointense in the portal and delayed phases.



Discussion

Principal Results

Diagnostic prediction of cancer by using data mining methods is an essential and significant application of EHRs [5]. From previous studies, features extracted from EHRs have proved to be the valid input of the diagnostic model [14,29]. In particular, the use of machine learning methods, especially deep learning methods for clinical information extraction, could facilitate in providing new evidences in computer-aided diagnosis. As the burden of liver cancer is widely accepted as one of the principal and universal challenges in health care and as patients with liver cancer are usually diagnosed at the terminal stage, the early and accurate diagnosis of liver cancer by radiology examination has great significance [30,31]. In contrast with previous studies of liver cancer diagnosis, our study focused on the identification of evidences for live cancer diagnosis from Chinese radiology reports. We selected APHE and PDPH as the known evidences for diagnosis according to the guidelines of CSCO. These 2 features were essential but not sufficient enough to represent the whole report and further be used to diagnose liver cancer. Furthermore, using FENLP, we also extracted uncertain numbers of radiological features from the report content, because we aimed to obtain new evidences for essential diagnosis. Therefore, the evidences for diagnosis were obtained both from clinical knowledge and the content of reports. For the recognition of APHE and PDPH, we applied BERT on word embedding in the deep learning method, which achieved state-of-the-art performance.

Word embedding is an essential step for sequencing labelling tasks. Previously popular models such as Word2Vec and Global Vector word representation focused on learning context-independent word representations. Recent advances in word representations based on language models, including ELMo, CoVe, and BERT, could dynamically improve the word representations and discriminate among multiple meanings of a word. In particular, based on the attention mechanism, BERT exhibited an upward trend and outperformed the previous models in many NLP tasks. Recognition of APHE and PDPH using traditional NLP methods had difficulties, because the related descriptions covered varied Chinese sentence structures and entity types (Table 1). For example, for hyperintense enhancement, the sentence pattern and phrase could have different styles due to the different writing habits of different radiologists or due to the use of Chinese abbreviations. Different from the Word2Vec model, BERT learned context-dependent word representations by using bidirectional transformers. The BiLSTM algorithms are widely used and easily implanted in sequence-related work such as entity extraction. We annotated all the characters in the Findings section manually with BIO tags and then applied the BERT-BiLSTM-CRF model to recognize APHE and PDPH. The high performance proved the feasibility of the BERT-BiLSTM-CRF model in information extraction from Chinese radiology reports.

In this study, among the recognition results of APHE and PDPH obtained by the 3 different models, the BERT-BILSTM-CRF model finally achieved the best performance for both APHE (F1 score 98.40%, precision 97.62%, and recall 99.19%) and PDPH (90.67%, 87.18%, and 94.44%, respectively) at the report level. For the 2 baseline models based on CRF, the model with a BiLSTM layer received a much higher F1 score than the model without a BiLSTM layer. The results indicated that, with the word embedding language model BERT and the BiLSTM model, the recognition of APHE and PDPH could result in much higher performance. To avoid the noise in the recognition results, we used the recognition results at the report level to be the input radiological features of the liver cancer diagnostic model. Report-level recognition concerned only continuous characters longer than 3 characters and specific Chinese abbreviations. Therefore, report-level results could represent whether the report contained the features of APHE or PDPH. The recognition of APHE and PDPH by BERT-BiLSTM-CRF was accurate enough to be the predictors of liver cancer diagnosis.

Only 2 fixed features of APHE and PDPH were not enough for liver cancer diagnosis. Therefore, we further performed the automatic NLP pipeline FENLP to extend the feature set based on Chinese grammar and radiological characteristics. Different from that of BERT-BILSTM-CRF, the number of features generated by FENLP was unknown and changed according to the training texts. In this study, we finally extracted 301 features. The top features were the typical morphology of the different locations, which were essential to the diagnosis of the liver diseases (Table S2 of Multimedia Appendix 1).

We chose the random forest as the liver cancer diagnostic model. With 2 kinds of features obtained by BERT-BILSTM-CRF and FENLP, random forest could reach an F1 score of 88.55%, which was much higher than the model using either kind of features. The performance of the diagnostic model using APHE and PDPH was slightly higher than that of the model using features extracted from FENLP. By contrast, FENLP produced much more features than BERT-BILSTM-CRF. We further ranked the features by the feature importance score computed by Gini impurity, which could reflect the degree of association with liver cancer. APHE and PDPH were the top 2 features with a clearly higher feature importance score compared with other features obtained by FENLP. The results indicated the strong association of APHE and PDPH with liver cancer, which coincided with the current clinical knowledge. Of the top features obtained by FENLP, the feature *high density of liver* had the highest feature importance score, which was the important and basic risk factor for the diagnosis of liver diseases. Broadening of hepatic fissures was an essential feature that existed in liver cirrhosis or in liver cancer progressed from liver cirrhosis [30]. Our results confirmed that the radiological features from FENLP could also be an evidence for diagnosis, which could further improve the diagnostic performance. Furthermore, the top features linked with liver cancer could extend the diagnostic evidence and be the supplementary features of APHE and PDPH.

Designing disease diagnostic models based on EHRs is a significantly important research field. Recently, NLP and deep learning-based models have been widely applied in many studies [7]. For instance, Sada et al designed and performed NLP-assisted radiology document classification for liver cancer detection. The model finally received an F1 score of 0.78 [23]. Compared with previous studies on clinical information extraction, the evidences for diagnosis in this study were identified based on the clinical knowledge from the guidelines of CSCO and the content of the reports. APHE and PDPH are 2 widely accepted evidences for disease diagnosis, and they have also proved to be essential features in our liver cancer diagnostic model. Other radiological features from FENLP enlarged the potential evidences for the diagnosis of liver cancer. Moreover, we utilized the BERT-BiLSTM-CRF model in this study, which achieved the state-of-the-art performance.

Limitations

Our study had the following limitations. The number of radiological features from FENLP was not fixed since all desirable features were retained, which might introduce some noise into the extracted radiological features. Besides, from the clinical knowledge in the guidelines of CSCO, we only extracted 2 characteristic features. In future, we will collect more evidences for diagnosis in order to further improve the performance and make the model more explanatory. Through the analysis of the misjudged samples in the recognition of APHE and PDPH, we identified the main error that occurred when the description of APHE and PDPH only included CT values. With the comparison of CT values in different phases, we could define these 2 features. However, our methods did not focus on the CT value extraction, and the number of these cases were small. In future studies, CT value extraction and analysis can avoid this kind of error and increase the prediction performance.

Conclusion

In this study, we developed a deep learning–based method for the recognition of evidences for disease diagnosis and designed a computer-aided liver cancer diagnosis framework. The diagnostic evidences contained APHE, PDPH, and radiological features extracted by FENLP. We proposed the BERT-based deep learning model BERT-BILSTM-CRF for recognizing the phrases of APHE and PDPH, which are the essential features associated with liver cancer diagnosis. Our work confirms that BERT-based deep learning model can be used and has desirable performance in the radiological feature extraction of Chinese radiology reports. Furthermore, this study was a comprehensive study for NLP and its application, focusing on Chinese radiology reports. The deep learning model proposed in this study for information extraction is expected to be extended to different types of Chinese clinical texts and other kinds of applications.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (No. 81701792 and No. 81971707) and the National Key Research and Development Program of China (No. 2018YFC0910404)

Authors' Contributions

HL proposed and designed the whole pipeline, analyzed the results, and wrote the paper. YX and ZY collected the original data, annotated the radiology reports, and provided clinical knowledge guidance in lexicon construction. ZZ, NW, and YH cleaned the data. RJ provided theoretical guidance and revised this paper. HC provided theoretical guidance and revised this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary data.
[DOCX File , 20 KB - jmir_v23i1e19689_app1.docx]

References

1. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
2. Wu S, Roberts K, Datta S, Du J, Ji Z, Si Y, et al. Deep learning in clinical natural language processing: a methodical review. J Am Med Inform Assoc 2020 Mar 01;27(3):457-470 [FREE Full text] [doi: 10.1093/jamia/ocz200] [Medline: 31794016]
3. Wu J, Liu X, Zhang X, He Z, Lv P. Master clinical medical knowledge at certificated-doctor-level with deep learning model. Nat Commun 2018 Oct 19;9(1):4352 [FREE Full text] [doi: 10.1038/s41467-018-06799-6] [Medline: 30341328]
4. Gálvez JA, Pappas JM, Ahumada L, Martin JN, Simpao AF, Rehman MA, et al. The use of natural language processing on pediatric diagnostic radiology reports in the electronic health record to identify deep venous thrombosis in children. J Thromb Thrombolysis 2017 Oct;44(3):281-290. [doi: 10.1007/s11239-017-1532-y] [Medline: 28815363]
5. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
6. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Inform 2019 May;125:37-46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.02.008] [Medline: 30914179]
7. Chen L, Song L, Shao Y, Li D, Ding K. Using natural language processing to extract clinically useful information from Chinese electronic medical records. Int J Med Inform 2019 Apr;124:6-12. [doi: 10.1016/j.ijmedinf.2019.01.004] [Medline: 30784428]
8. Li P, Yuan Z, Tu W, Yu K, Lu D. Medical Knowledge Extraction and Analysis from Electronic Medical Records Using Deep Learning. Chin Med Sci J 2019 Jun 30;34(2):133-139. [doi: 10.24920/003589] [Medline: 31315754]
9. Mikolov T, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. In: Advances in Neural Information Processing Systems 26 (NIPS 2013). United States: NIPS Foundation; 2013:3111-3119.
10. Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation. 2014 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP); 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
11. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K. Deep contextualized word representations. 2018 Presented at: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2018; New Orleans, LA,USA p. 2227-2237 URL: https://arxiv.org/abs/1802.05365
12. Devlin J, Chang M, Kristina K, Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2019; Minneapolis, MN, USA p. 4171-4186. [doi: 10.18653/v1/N19-1423]
13. Zhiheng H, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. 2015. URL: https://arxiv.org/abs/1508.01991 [accessed 2020-01-03]
14. Zhang X, Zhang Y, Zhang Q, Ren Y, Qiu T, Ma J, et al. Extracting comprehensive clinical information for breast cancer using deep learning methods. Int J Med Inform 2019 Dec;132:103985. [doi: 10.1016/j.ijmedinf.2019.103985] [Medline: 31627032]
15. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019 Mar;25(3):433-438. [doi: 10.1038/s41591-018-0335-9] [Medline: 30742121]
16. Wang Q, Zhou Y, Ruan T, Gao D, Xia Y, He P. Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition. J Biomed Inform 2019 Apr;92:103133 [FREE Full text] [doi: 10.1016/j.jbi.2019.103133] [Medline: 30818005]
17. Ji B, Li S, Yu J, Ma J, Tang J, Wu Q, et al. Research on Chinese medical named entity recognition based on collaborative cooperation of multiple neural network models. J Biomed Inform 2020 Apr;104:103395. [doi: 10.1016/j.jbi.2020.103395] [Medline: 32109551]
18. Bozkurt S, Alkim E, Banerjee I, Rubin DL. Automated Detection of Measurements and Their Descriptors in Radiology Reports Using a Hybrid Natural Language Processing Algorithm. J Digit Imaging 2019 Aug;32(4):544-553 [FREE Full text] [doi: 10.1007/s10278-019-00237-9] [Medline: 31222557]
19. Steinkamp JM, Chambers C, Lalevic D, Zafar HM, Cook TS. Toward Complete Structured Information Extraction from Radiology Reports Using Machine Learning. J Digit Imaging 2019 Aug;32(4):554-564 [FREE Full text] [doi: 10.1007/s10278-019-00234-y] [Medline: 31218554]
20. Flynn RWV, Macdonald TM, Schembri N, Murray GD, Doney ASF. Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes. Pharmacoepidemiol Drug Saf 2010 Aug;19(8):843-847. [doi: 10.1002/pds.1981] [Medline: 20602346]

21. Yetisgen-Yildiz M, Gunn ML, Xia F, Payne TH. A text processing pipeline to extract recommendations from radiology reports. J Biomed Inform 2013 Apr;46(2):354-362 [FREE Full text] [doi: 10.1016/j.jbi.2012.12.005] [Medline: 23354284]
22. Hassanpour S, Bay G, Langlotz CP. Characterization of Change and Significance for Clinical Findings in Radiology Reports Through Natural Language Processing. J Digit Imaging 2017 Jun;30(3):314-322 [FREE Full text] [doi: 10.1007/s10278-016-9931-8] [Medline: 28050714]
23. Sada Y, Hou J, Richardson P, El-Serag H, Davila J. Validation of Case Finding Algorithms for Hepatocellular Cancer From Administrative Data and Electronic Health Records Using Natural Language Processing. Med Care 2016 Feb;54(2):e9-14 [FREE Full text] [doi: 10.1097/MLR.0b013e3182a30373] [Medline: 23929403]
24. Xu H, Fu Z, Shah A, Chen Y, Peterson NB, Chen Q, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. AMIA Annu Symp Proc 2011;2011:1564-1572 [FREE Full text] [Medline: 22195222]
25. Guidelines of Chinese Society of Clinical Oncology (CSCO): Hepatocellular Carcinoma. Beijing,China: Chinese Society of Clinical Oncology; 2018.
26. Liu H, Xu Y, Zhang Z, Wang N, Huang Y, Hu Y, et al. A Natural Language Processing Pipeline of Chinese Free-text Radiology Reports for Liver Cancer Diagnosis. IEEE Access 2020 Aug 28;8:159110-159119. [doi: 10.1109/ACCESS.2020.3020138]
27. Tawfik NS, Spruit MR. Evaluating sentence representations for biomedical text: Methods and experimental results. J Biomed Inform 2020 Apr;104:103396. [doi: 10.1016/j.jbi.2020.103396] [Medline: 32147441]
28. GitHub. URL: https://github.com/google-research/bert [accessed 2020-03-01]
29. Pathak S, van Rossen J, Vijlbrief O, Geerdink J, Seifert C, van Keulen M. Post-Structuring Radiology Reports of Breast Cancer Patients for Clinical Quality Assurance. IEEE/ACM Trans Comput Biol Bioinform 2020;17(6):1883-1894. [doi: 10.1109/TCBB.2019.2914678] [Medline: 31059453]
30. Kudo M, Trevisani F, Abou-Alfa GK, Rimassa L. Hepatocellular Carcinoma: Therapeutic Guidelines and Medical Treatment. Liver Cancer 2016 Nov;6(1):16-26 [FREE Full text] [doi: 10.1159/000449343] [Medline: 27995084]
31. Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, WHO Classification of Tumours Editorial Board. The 2019 WHO classification of tumours of the digestive system. Histopathology 2020 Jan;76(2):182-188 [FREE Full text] [doi: 10.1111/his.13975] [Medline: 31433515]

Abbreviations

APHE: hyperintense enhancement in the arterial phase
BERT: Bidirectional Encoder Representations from Transformers
BiLSTM: bidirectional long short-term memory
BIO: Begin, Inside, Outside
CRF: conditional random field
CSCO: Chinese Society of Clinical Oncology
CT: computed tomography
EHR: electronic health record
ELMo: Embeddings from Language Model
FENLP: feature extraction using the rule-based natural language processing
NER: named-entity recognition
NLP: natural language processing
PDPH: hypointense in the portal and delayed phases

Edited by G Eysenbach; submitted 29.04.20; peer-reviewed by M Torii, J Zheng; comments to author 01.06.20; revised version received 30.06.20; accepted 11.11.20; published 12.01.21.

Please cite as:
Liu H, Zhang Z, Xu Y, Wang N, Huang Y, Yang Z, Jiang R, Chen H
Use of BERT (Bidirectional Encoder Representations from Transformers)-Based Deep Learning Method for Extracting Evidences in Chinese Radiology Reports: Development of a Computer-Aided Liver Cancer Diagnosis Framework
J Med Internet Res 2021;23(1):e19689
URL: http://www.jmir.org/2021/1/e19689/
doi:10.2196/19689
PMID:33433395

Original Paper

Disease Concept-Embedding Based on the Self-Supervised Method for Medical Information Extraction from Electronic Health Records and Disease Retrieval: Algorithm Development and Validation Study

Yen-Pin Chen[1,2,3], MD; Yuan-Hsun Lo[4], PhD; Feipei Lai[1], PhD; Chien-Hua Huang[3], MD, PhD

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei City, Taiwan
[2]Department of Emergency Medicine, National Taiwan University BioMedical Park Hospital, Hsinchu County, Taiwan
[3]Department of Emergency Medicine, National Taiwan University Hospital, Taipei City, Taiwan
[4]Department of Applied Mathematics, National Pingtung University, Pingtung City, Taiwan

Corresponding Author:
Chien-Hua Huang, MD, PhD
Department of Emergency Medicine
National Taiwan University Hospital
No 7, Chung Shan S Rd (Zhongshan S Rd), Zhongzheng District
Taipei City
Taiwan
Phone: 886 2 2312 3456
Email: chhuang5940@ntu.edu.tw

Abstract

Background: The electronic health record (EHR) contains a wealth of medical information. An organized EHR can greatly help doctors treat patients. In some cases, only limited patient information is collected to help doctors make treatment decisions. Because EHRs can serve as a reference for this limited information, doctors' treatment capabilities can be enhanced. Natural language processing and deep learning methods can help organize and translate EHR information into medical knowledge and experience.

Objective: In this study, we aimed to create a model to extract concept embeddings from EHRs for disease pattern retrieval and further classification tasks.

Methods: We collected 1,040,989 emergency department visits from the National Taiwan University Hospital Integrated Medical Database and 305,897 samples from the National Hospital and Ambulatory Medical Care Survey Emergency Department data. After data cleansing and preprocessing, the data sets were divided into training, validation, and test sets. We proposed a Transformer-based model to embed EHRs and used Bidirectional Encoder Representations from Transformers (BERT) to extract features from free text and concatenate features with structural data as input to our proposed model. Then, Deep InfoMax (DIM) and Simple Contrastive Learning of Visual Representations (SimCLR) were used for the unsupervised embedding of the disease concept. The pretrained disease concept-embedding model, named EDisease, was further finetuned to adapt to the critical care outcome prediction task. We evaluated the performance of embedding using t-distributed stochastic neighbor embedding (t-SNE) to perform dimension reduction for visualization. The performance of the finetuned predictive model was evaluated against published models using the area under the receiver operating characteristic (AUROC).

Results: The performance of our model on the outcome prediction had the highest AUROC of 0.876. In the ablation study, the use of a smaller data set or fewer unsupervised methods for pretraining deteriorated the prediction performance. The AUROCs were 0.857, 0.870, and 0.868 for the model without pretraining, the model pretrained by only SimCLR, and the model pretrained by only DIM, respectively. On the smaller finetuning set, the AUROC was 0.815 for the proposed model.

Conclusions: Through contrastive learning methods, disease concepts can be embedded meaningfully. Moreover, these methods can be used for disease retrieval tasks to enhance clinical practice capabilities. The disease concept model is also suitable as a pretrained model for subsequent prediction tasks.

(J Med Internet Res 2021;23(1):e25113) doi:10.2196/25113

KEYWORDS

electronic health record; EHR; disease embedding; disease retrieval; emergency department; concept; extraction; deep learning; machine learning; natural language processing; NLP

Introduction

Background

Diagnosing a disease is like putting together a puzzle. When more pieces match, we can decipher the picture more easily. Or, if we have a reference picture of the puzzle, matching the pieces can be accomplished more easily. Many "puzzles" have been collected in electronic health records (EHRs), which contain abundant information about patients and diseases and represent a treasure trove for medical research. However, EHRs are challenging to use effectively due to the heterogeneity of the data types they can contain [1-3].

Sometimes, doctors in emergency departments must develop treatment plans based on limited information from seriously ill patients. In such cases, only demographic information, vital signs (eg, blood pressure, respiratory rate, and blood oxygen saturation), and major discomfort information can be obtained. However, a patient's medical history may be collected through old records in the hospital database, and these records can provide doctors with vital information that can help them to identify serious diseases [4]. Based on the doctor's knowledge and experience, the patient's information can characterize the disease, enable the doctor to diagnose the patient, and allow the doctor to develop an appropriate treatment plan to prevent the disease from worsening.

Doctors can make great use of the rich information contained within patients' EHRs. However, it is almost impossible for EHRs to be of help to doctors if they are not organized. Full use of the medical information contained within EHRs in clinical practice greatly enhances doctors' abilities to treat patients [3,5,6].

The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) code [7] is a well-established classification system for diseases that contains approximately 70,000 diagnoses. These diagnoses are highly specific and contain meaningful words that can be used for patient medical records. ICD codes can easily be used as keywords to search for an EHR. However, it is difficult to search for records without the ICD code of a patient.

With the rapid development of natural language processing and deep learning during this decade, several models have been released to manage EHRs [1,8-10] using several primary tasks, including disease classification, prediction of clinical events, and concept embedding [2]. Due to the lack of gold standard label data for model training, concept embedding is critical to all EHR-related tasks [1,2,11,12]. Concept embedding can be regarded as a pretrained model that can be finetuned with limited label data for further event prediction, disease classification, and detection, including the diagnosis of diseases. Concept embedding can also be used for data retrieval [13,14], and can be further integrated into medical software, web services, and apps used by patients for disease screening purposes.

Therefore, the concept-embedding model is essential to EHR research because it can organize rich medical experience and knowledge. We aim to develop a disease embedding model that can cluster disease patterns based on initial data (eg, demographic information, vital signs, chief complaint, and medical history) collected from patients who have just arrived at a hospital.

Related Works

Structural data (eg, demographic information, blood pressure, and heart rate) and free text records are collected in an EHR. In traditional medical research, free text records are more difficult to analyze, and significant human resources are required to classify free text into tables using rule-based methods [1,2,10]. With the vigorous development of deep learning, several language models for extracting features from free text records have become available [15-17]. Long short-term memory [18] is a model architecture that can feed sequence data and properly ignore input or memory history data to extract accurate features. It was widely used on natural language processing tasks before Transformer [19] was proposed.

Transformer contains the multihead self-attention model, which can learn the attention (attention can be regarded as relevance) of words within the text. Transformer was released to perform translation tasks and speed up learning by feeding all data simultaneously (instead of feeding them one-by-one as is the case with long short-term memory). Bidirectional Encoder Representations from Transformers (BERT) [16] is an extension of Transformer, which has achieved excellent performance ratings in several competitions. The first version of BERT contained only Transformer's encoder, which pretrained the embedding of subwords (words broken into pieces) unsupervised through a cloze procedure and subsequent sentence prediction. Because translation was the original goal of Transformer, the multilingual pretrained BERT is suitable for EHR research in Taiwan because Taiwanese medical records contain English and Chinese.

Previous methods for concept embedding of the EHR model include long short-term memory, convolutional neural networks, and autoencoders [1,2,11,12]. An autoencoder is a widely used unsupervised method for concept embedding. It contains an encoder that converts the original input data into a hidden embedding vector, while the decoder reconstructs a result similar to the input data from the hidden embedding vector. In the EHR field, autoencoders outperform traditional machine learning methods [1]. However, the risk of reidentification is a critical issue for autoencoders. Some studies exist on protecting patient privacy with autoencoders [20,21]. However, if the autoencoder is not designed to protect privacy, there is still a risk of reidentification because the decoder is designed to minimize the difference between the original data and the output.

In the EHR field, the contrastive self-supervised learning methods that belong to unsupervised learning, including Deep InfoMax (DIM) [22] and Simple Contrastive Learning of Visual

Representations (SimCLR) [23], may be superior. DIM adopted discriminators (replacing the autoencoder's decoder) that maximize mutual information between the input data and the embedding vectors. SimCLR minimizes the embedding distance within groups while maximizing the embedding distance among groups. Compared with autoencoders, contrastive self-supervised learning methods demonstrate superior results when performing classification tasks [22], even at the level of supervised learning [23], but no decoder model produces the risk of reidentification.

Due to the vectorization of objects of interest, disease concept-embedding can be used for retrieval [24]. Vector representation of big data can be operated simultaneously using linear algebra methods [25]. EHR-based clinical information retrieval [3,13,26,27] has been proposed. One study [13] focused on creating test sets for information retrieval research. Two studies [3,26] used keyword-based information retrieval systems, which did not use the semantic analysis method and could not identify negation semantics. Another study [27]developed a document-level semantic-based query recommendation search engine. Furthermore, these studies only focused on EHR-related retrieval methods that cannot be directly applied to predictive models. No objective score or standard exists to evaluate the performance of the concept-embedding model or the information retrieval [3]. However, visualization is a widely used method for evaluating the performance of concept embedding [10]. The t-distributed stochastic neighbor embedding (t-SNE) [28] method reduces the embedding's dimensionality to visualize the clustering effects of unsupervised learning. Although not objective, another evaluation method is sampling each cluster group and determining its topic [13,29]. Therefore, we can use a patient's data to retrieve neighbor samples and verify the consistency of the samples' concepts. Furthermore, if a new patient data is provided, concept embedding can be used to retrieve the top five most similar data, after which we can evaluate the retrieval capabilities.

In some papers, subsequent applications of classification have been adopted [2]. For emergency department employees, a patient's condition severity and emergency information are most relevant. We introduced a label for critical care outcome, defined as either (1) an intensive care unit admission or (2) death in the hospital after an emergency department visit [30-35]. This outcome was used to finetune the pretrained embedding model for the classification task. Furthermore, we compared the performance of the embedding model with that of state-of-the-art models.

Objective

This study aims to create a model that uses limited patient information to extract disease concept-embedding for disease retrieval and classification tasks. Accordingly, we present a disease concept-embedding method, the EDisease model, based on Transformer and the contrastive self-supervised learning method.

Methods

Materials

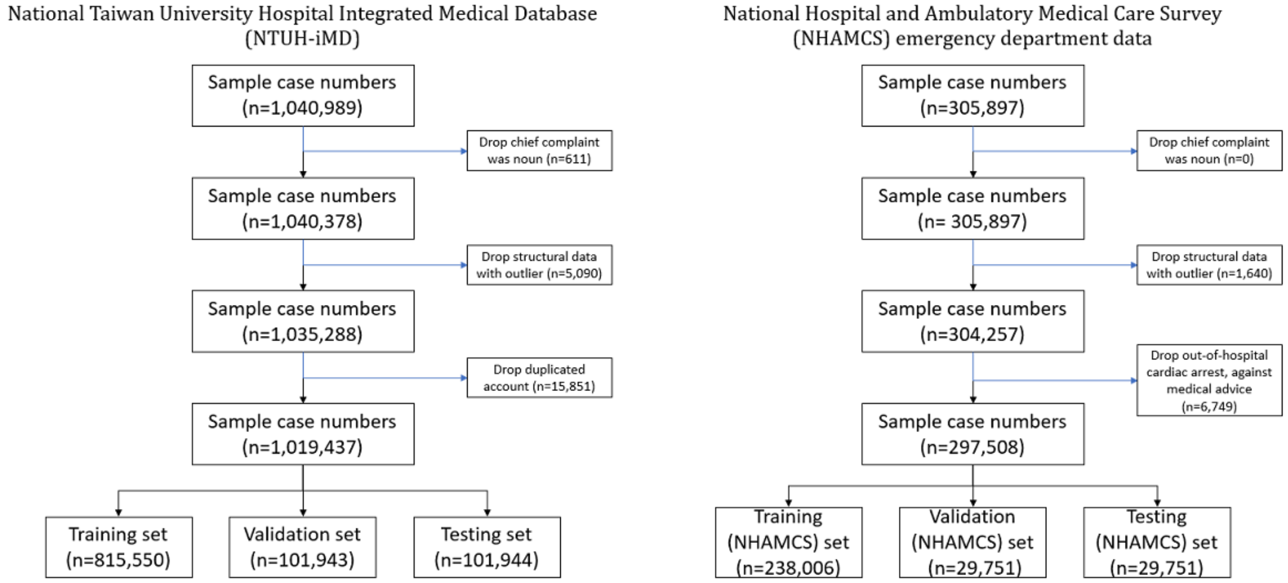
A sample of 1,040,989 emergency department visits was collected from the National Taiwan University Hospital Integrated Medical Database, a private EHR data set. The sample included the patients of National Taiwan University Hospital (one medical center in Taipei City and two regional hospitals in Hsinchu City and Yunlin County), who visited the hospital from January 1, 2013 to December 31, 2017. This study was approved by Research Ethics Committee B at National Taiwan University Hospital (201902078RINB).

In the data preprocessing stage, if the chief complaint was missing data or the structural data were unreasonable (eg, blood pressure was higher than 300 mmHg, diastolic blood pressure was higher than systolic blood pressure, heart rate was higher than 250 beats/minute, respiration rate was higher than 100 times/minute, body temperature was higher than 48°C or lower than 20°C, and body weight and height were higher than 400kgs and 250cms, respectively), the samples were discarded. In this stage, structural data with missing values were retained.

Individual patients may have visited the hospital several times during the specified period, and each visit corresponded to one account. Thus, each patient's ID number may have been associated with multiple EHRs. In the National Taiwan University Hospital, if a patient visits more than once within 24 hours, the last account number can be retained for reuse using another revisit flag. In this study, we retained only the last sample with the same account but different revisit flags. Furthermore, for each ID number, there are several medical records in the EHR, and we used the triage time to filter future records.

We then split the sample of 1,019,437 visits into sizes of 815,550 (80.0%), 101,943 (10.0%), and 101,944 (10.0%) for the training, validation, and test sets, respectively (Figure 1).

Figure 1. Data analysis flowcharts.



We also collected National Hospital and Ambulatory Medical Care Survey emergency department data from 2007 to 2017 [36]. This survey data is available in a public EHR database provided annually by the Centers for Disease Control and Prevention National Center for Health Statistics. We preprocessed the structured data using the same preprocessing method described previously and excluded patients with death on emergency department arrival, or those who left before being seen by a practitioner or against medical advice. Structured chief complaints and comorbidities were converted into free text as the chief complaints and past medical history, respectively. For the National Hospital and Ambulatory Medical Care Survey data, we split the 297,508 visit sample into sizes of 238,006 (80.0%), 29,751 (10.0%), and 29,751 (10.0%) for training, validation, and test sets, respectively (Figure 1).

The means and standard deviations were counted after discarding the missing data. Moreover, critical care outcome was selected as the outcome. Critical care outcome refers to the union of intensive care unit admission and death. We distinguished the definitions of intensive care unit admission and death for each data set. In the National Taiwan University Hospital Integrated Medical Database, intensive care unit admission is defined as whether an intensive care unit record existed within three days after a patient's emergency department visit. Death is defined as whether a death record existed within three days after a patient's emergency department visit [30,31]. In the National Hospital and Ambulatory Medical Care Survey, intensive care unit admission is defined as direct admission to an intensive care unit, and death is defined as in-hospital death (Table 1).

Emergency staff work in shifts to treat emergency patients promptly. Therefore, an alert system for impending emergency events is useful. Consequently, we chose a three-day interval as the cut-off point for disease progression, resulting in the outcome in the National Taiwan University Hospital Integrated Medical Database being defined as any critical event within three days after an emergency department visit. In contrast, our definition of critical care outcome in the National Hospital and Ambulatory Medical Care Survey was identical to that of the reference [32-34].

Table 1. Structural information and outcomes.

Variable	National Taiwan University Hospital Integrated Medical Database				NHAMCS[a]
	Medical center (Taipei City) (n=541,366)	Regional hospitals (Hsinchu City) (n=233,921)	Regional hospitals (Yunlin County) (n=244,150)	Total (N=1,019,437)	Emergency department data (n=297,508)
Age, mean years (SD)	43.1 (27.0)	42.2 (27.2)	46.9 (26.0)	43.7 (26.9)	37.4 (24.0)
Gender (male), n (%)	266,891 (49.3)	118,815 (50.8)	129,221 (52.9)	514,927 (50.5)	135,147 (45.4)
Vital signs					
Systolic blood pressure, mmHg (SD)	133.0 (26.9)	140.2 (28.5)	136.2 (26.2)	135.4 (27.3)	132.9 (23.5)
Diastolic blood pressure, mmHg (SD)	77.8 (15.2)	77.7 (15.4)	80.3 (17.2)	78.3 (15.8)	77.7 (14.6)
Heart Rate, BPM[b] (SD)	97.6 (26.8)	97.0 (26.0)	90.0 (20.8)	95.8 (25.6)	91.1 (22.8)
Saturation, % SpO_2 (SD)	97.0 (2.8)	97.1 (3.0)	97.6 (5.6)	97.2 (3.64)	97.3 (6.2)
Respiratory rate, BrPM[c] (SD)	19.9 (3.7)	20.4 (2.5)	17.6 (2.4)	19.5 (3.4)	19.3 (4.4)
Body temperature, °C (SD)	37.1 (0.9)	37.1 (1.0)	36.9 (1.0)	37.1 (1.0)	36.8 (0.6)
Body height, cm (SD)	151.6 (27.9)	153.4 (25.5)	157.4 (20.4)	152.8 (26.6)	N/A[d]
Body weight, kg (SD)	53.3 (22.3)	47.0 (25.1)	54.4 (21.9)	52.5 (22.8)	N/A
Pain index, value (SD)	2.6 (3.2)	3.1 (3.3)	2.3 (2.9)	2.7 (3.2)	4.8 (3.7)
Glasgow coma scale, value (SD)	14.8 (1.3)	14.7 (1.4)	14.6 (1.7)	14.7 (1.4)	14.5 (2.1)
Eye response, value (SD)	4.0 (0.3)	4.0 (0.3)	3.9 (0.4)	4.0 (0.3)	N/A
Verbal response, value (SD)	4.9 (0.5)	4.9 (0.6)	4.8 (0.7)	4.9 (0.6)	N/A
Motor response, value (SD)	5.9 (0.4)	5.9 (0.5)	5.9 (0.7)	5.9 (0.5)	N/A
Outcomes					
Admission to ICU, n (%)	1,958 (0.4)	336 (0.1)	613 (0.3)	2,907 (0.2)	4,375 (1.5)
Death, n (%)	1,993 (0.4)	706 (0.3)	891 (0.4)	3,590 (0.4)	381 (0.1)
Critical care outcome, n (%)	3,906 (0.7)	1,039 (0.4)	1,492 (0.6)	6,437 (0.6)	4,755 (1.6)

[a] NHAMCS: National Hospital and Ambulatory Medical Care Survey.

[b] BPM: beats per minute.

[c] BrPM: breaths per minute.

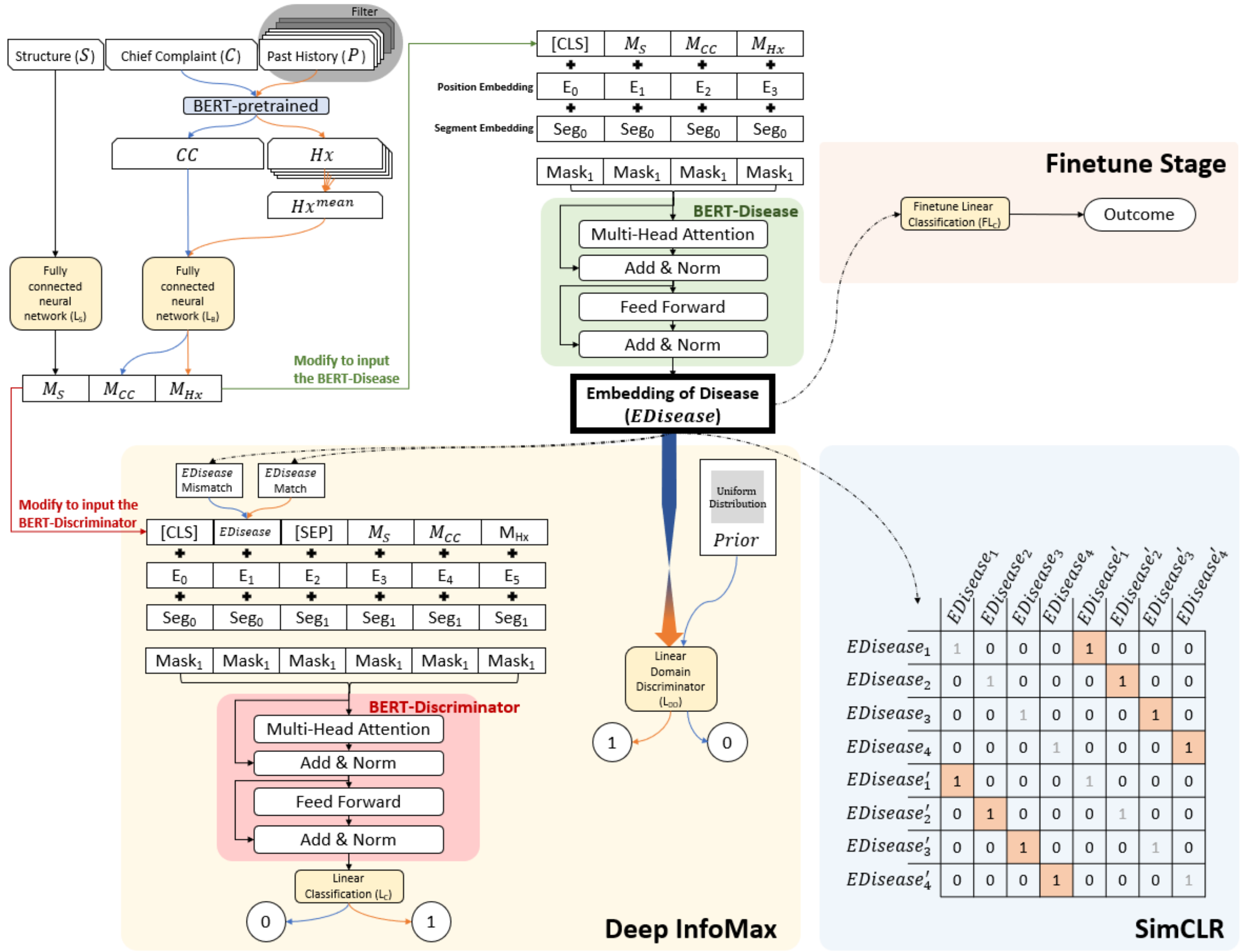
[d] N/A: not available.

Model Architecture

Transformer was adopted as the central architecture in this study, based on BERT with the PyTorch adaptation released by the HuggingFace team [37]. Initially, we collected structural data $S \in \mathbf{R}^{15}$, including age, gender, blood pressure, heart rate, blood oxygen saturation, respiratory rate, body temperature, height, weight, subjective pain score, and Glasgow coma scale. We filled each unavailable data point with the average value of each S and performed noise processing on each element in S with a random scale based on the standard deviation of each element [38]. The fully-connected neural network for structural data L_S is a multilayer perceptron, which converts S to $M_S \in \mathbf{R}^{96}$ (Figure 2). For the free text medical record data, the pretrained "bert-base-multilingual-cased" model (ie, *BERT*-pretrained) is used to extract features CC, Hx, $\in \mathbf{R}^{768}$ from the free text (chief complaint, C, and past medical history, P). Because the number of past medical history data points changes, we averaged the features of all Hx and obtain the Hx^{mean} that represents the patient's medical history. If there is no medical history in the EHR, we use a padding vector to fill the Hx^{mean}. The fully-connected neural network L_B for $BERT$ is a multilayer perceptron that converts CC and Hx^{mean} to M_{cc} and $M_{Hx} \in \mathbf{R}^{96}$, respectively.

n is the total number of samples, k denotes the numbers of medical history.

Figure 2. Deep-learning model architecture.



	$EDisease_1$	$EDisease_2$	$EDisease_3$	$EDisease_4$	$EDisease'_1$	$EDisease'_2$	$EDisease'_3$	$EDisease'_4$
$EDisease_1$	1	0	0	0	1	0	0	0
$EDisease_2$	0	1	0	0	0	1	0	0
$EDisease_3$	0	0	1	0	0	0	1	0
$EDisease_4$	0	0	0	1	0	0	0	1
$EDisease'_1$	1	0	0	0	1	0	0	0
$EDisease'_2$	0	1	0	0	0	1	0	0
$EDisease'_3$	0	0	1	0	0	0	1	0
$EDisease'_4$	0	0	0	1	0	0	0	1

We embed the diseases to concepts by adopting another BERT architecture as *BERT*-Disease (β_{ED}), a 12-layer Transformer encoder with 12 self-attention heads and a hidden size of 96. We concatenate M_S, M_{cc}, M_{Hx} and modify it with a leading head [CLS] as the input embedding. Position embedding indicates each element's position in the input embedding sequence, and segment embedding indicates the level to which the element belongs [16]. All segment embeddings are of the same level at this stage, but we reserved this settable variable for the subsequent work in this study. We obtain the sum of input embedding, position embedding, and segment embedding as the input of β_{ED}.

The *Mask* ∈ {0,1} indicates whether the element is valid (1) or invalid (0). For example, some patients did not have any medical history records in their EHRs. In these cases, the padding Hx^{mean} will be used as the input; however, the padding element should be ignored. Furthermore, when we gathered more information from some patients, *Mask* enabled us to ignore other padding elements and is, therefore, essential to this extensible model. We also used the random mask on M_{Hx} for the augmentation of *M* to improve the ability to embed the disease concept in a state with less information. This augmentation of a random history mask is also adopted in the subsequent work with another modification *M*. In this part, we obtained the embedding of disease:

DIM was adopted to further construct accurate disease embedding. DIM's core principle is to maximize the mutual information between the samples and embedding [22]. The mutual information on *M* and *EDisease* is as follows, where *p(m)* denotes the probability of each visit in the sample.

We assumed that each visit was unique. Therefore, $p(m), \forall m \in M$ are the same constant. Moreover, we assumed the number of disease embedding data points was limited and smaller than the number of visits, such that . The result is that $B_{ED}: M \rightarrow EDisease$ is a surjective function also known as onto, so we then obtain $p(\varepsilon) = p(\varepsilon|M)p(M)$.

We then set the function *DIM Loss* as follows and minimized it to maximize the mutual information.

We achieved this goal by adopting another BERT architecture, *BERT-Discriminator* (β_D), which shares the same model architecture as B_{ED} and estimate $\beta_D(\varepsilon, m) = p(\varepsilon|m)$. In this stage, *M* is modified not only by [CLS] but also by *EDisease* and [SEP]. The [SEP] is used to separate *EDisease* and *M* and this modification is similar to “the next sentence prediction” task in BERT [16]. We used segment embedding with a level of 0 to modify the [CLS] and *EDisease* with a level of 1 to modify [SEP] and *M*. We set $Y_{(\varepsilon,m)}=1$, when $i=j$, $\forall m_i,\varepsilon_j$ and $Y_{(\varepsilon,m)}=0$, when $i\neq j$, $\forall m_i$, ε_j and simplified as follows.

A part of *Loss* could now be discovered using the cross-entropy loss, similar to noise-contrastive estimation [39,40]. However, one part remained to be solved. We had to establish an adequate prior distribution to constrain the embedding distribution as proposed by DIM [22] and to minimize the $\sum p(\varepsilon)\log(p(\varepsilon))$.

The function $F(x) = x \log x$ is convex [41]. Then,

and two parts are equal if and only if every number x_i is the same (it can be proved by the Lagrange multiplier), which is equivalent to the uniform distribution that can minimize *Loss*. Consequently, we adopted a uniform distribution *Prior* on $[–1,1]^{96}$ to constrain $p(\varepsilon)$.

Generative adversarial networks [42] can be used to constrain an embedded distribution to a prior distribution [22]. The final distribution of EDisease is difficult to distinguish from the prior distribution uniform distribution using a linear domain discriminator L_{DD}. This discriminator detects whether each input distribution belongs to the prior or EDisease domain and estimates the distribution divergence of each domain. Moreover, this goal can be achieved using the generative adversarial network iteration algorithm.

Therefore, we simplify our goal to the following equation:

For a more robust concept embedding, we also introduced the SimCLR loss. Just as with the augmentation in SimCLR, each m_i is augmented to another m'_i by random noise. Ideally, the cosine similarity of $B_{ED}(m_i)$ to $B_{ED}(m'_i)$ is 1 (positive sample), and the cosine similarities of $B_{ED}(m_i)$ to $B_{ED}(m_j)$ and $B_{ED}(m'_j)$ are -1 (negative sample), $\forall j \neq i$.

Therefore, the final loss function is:

$$Loss = DIM\ Loss + SimCLR\ Loss$$

In the further finetuning stage for predicting the outcome, EDisease is fed to a multilayer linear model FL_C (Figure 2).

We can add an extended model for the extended information to enhance the architecture (see Multimedia Appendix 1). We demonstrated the extensible model by obtaining M_{PI} after feeding an illness to the model and concatenating it to the vanilla *M*. The subsequent training methods are identical to those in Figure 2.

For improved performance with free text feature extraction, we again pretrained *BERT-pretrained* using the free text medical records in the National Taiwan University Hospital Integrated Medical Database. In the medical record, the chief complaint in triage was the patient’s words for their discomfort, and present illness, *I*, was the doctor’s word for the patient’s problem; thus, the similarity between the chief complaint and patient illness should, ideally, be high. Furthermore, some patients had several medical history records, and their subsets may also have been highly similar. Consequently, CC, PI, Hx_1^{mean}, $Hx_2^{mean} \in \mathbf{R}^{768}$ were extracted by *BERT-pretrained*, the pair CC, PI representing the same patient’s complaints, and the pair Hx_1^{mean}, Hx_2^{mean} representing the same patient’s medical history. We then used the contrastive learning method used previously to pretrain *BERT-pretrained*. Due to graphic processing unit memory limitations, after again pretraining the language model, we fixed the model’s weights such that the grading of subsequent training would not change the weight of the model.

We used Adam optimization [43] for the hyperparameters and set the learning rate as $1\times10^{–4}$, with a minibatch size of 1024. The hyperparameter γ=0.1 was chosen in this study. The source code is available on GitHub [44].

Evaluation

We selected the training set and fed it into the model to obtain the set of disease embedding. We then used the t-SNE to reduce the dimensionality of the embedding to two for visualization. We enhanced visualization based on demographic information, hospital level, triage level, whether medical history existed, and outcome. Furthermore, we illustrated the model’s ability to cluster diseases by selecting patients from the validation set as queries to retrieve similar patients in the training set. The relevance of each result to the query was judged by a seasoned doctor, who used the number of hits on the top 5 results to score from 0-5.

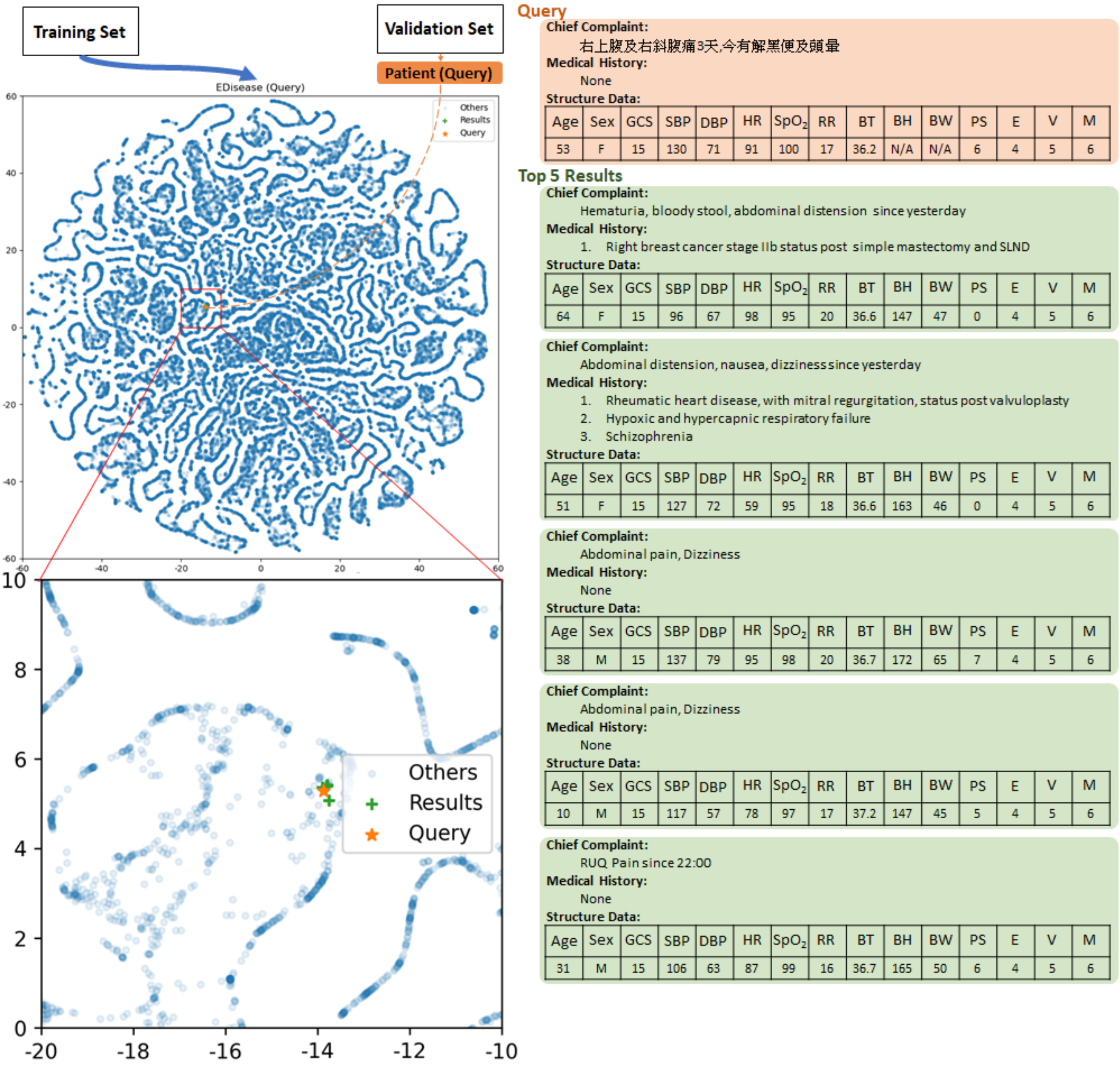
Furthermore, we finetuned the model based on the outcome and compared it with the published model based on the receiver operating characteristic curve. In order to check whether the embedding could be used as a pretrained model, even if only a small amount of label data was collected, we also took 10,000 samples from the validation (National Hospital and Ambulatory Medical Care Survey) set as a small data set for finetuning. Then, we performed ablation studies by deleting each unsupervised learning method and comparing them.

Results

We used the t-SNE method to reduce the 96-dimensional embedding to 2-dimensional embedding and enhance the visualization by gender, age (10-year intervals), hospital level, triage level, existence of a medical history, and outcome (see Multimedia Appendix 2).

We sampled the patients in the validation set as queries to retrieve the 5 most similar patients in the training set in the embedding space. Moreover, a doctor evaluated the relevance of 25 random query results (see Multimedia Appendix 3, Figure 3, and Multimedia Appendix 4).

Figure 3. Disease Retrieval Demonstration. The query subject (orange) was a 53-year-old female patient who suffered from abdominal pain in the upper-right quadrant to right flanks for three days and noticed dizziness and tarry stool on the day of the interview. Through the retrieval, we obtained the top five similar patients (green) whose symptoms were hematuria, bloody stools, abdominal distention, abdominal pain, dizziness, and abdominal pain in the upper-right quadrant. GCS: Glasgow coma scale. SBP: Systolic blood pressure. DBP: Diastolic blood pressure. HR: Heart rate. SpO$_2$: Blood oxygen saturation. RR: Respiratory rate. BT: Body temperature in Celsius. BH: Body height. BW: Body weight. PS: Pain scale. E: Eye response in Glasgow coma scale. V: Verbal response in Glasgow coma scale. M: Motor response in Glasgow coma scale. N/A : Not available.



Query

Chief Complaint:

右上腹及右斜腹痛3天,今有解黑便及頭暈

Medical History:

None

Structure Data:

Age	Sex	GCS	SBP	DBP	HR	SpO$_2$	RR	BT	BH	BW	PS	E	V	M
53	F	15	130	71	91	100	17	36.2	N/A	N/A	6	4	5	6

Top 5 Results

Chief Complaint:

Hematuria, bloody stool, abdominal distension since yesterday

Medical History:

1. Right breast cancer stage IIb status post simple mastectomy and SLND

Structure Data:

Age	Sex	GCS	SBP	DBP	HR	SpO$_2$	RR	BT	BH	BW	PS	E	V	M
64	F	15	96	67	98	95	20	36.6	147	47	0	4	5	6

Chief Complaint:

Abdominal distension, nausea, dizziness since yesterday

Medical History:

1. Rheumatic heart disease, with mitral regurgitation, status post valvuloplasty
2. Hypoxic and hypercapnic respiratory failure
3. Schizophrenia

Structure Data:

Age	Sex	GCS	SBP	DBP	HR	SpO$_2$	RR	BT	BH	BW	PS	E	V	M
51	F	15	127	72	59	95	18	36.6	163	46	0	4	5	6

Chief Complaint:

Abdominal pain, Dizziness

Medical History:

None

Structure Data:

Age	Sex	GCS	SBP	DBP	HR	SpO$_2$	RR	BT	BH	BW	PS	E	V	M
38	M	15	137	79	95	98	20	36.7	172	65	7	4	5	6

Chief Complaint:

Abdominal pain, Dizziness

Medical History:

None

Structure Data:

Age	Sex	GCS	SBP	DBP	HR	SpO$_2$	RR	BT	BH	BW	PS	E	V	M
10	M	15	117	57	78	97	17	37.2	147	45	5	4	5	6

Chief Complaint:

RUQ Pain since 22:00

Medical History:

None

Structure Data:

Age	Sex	GCS	SBP	DBP	HR	SpO$_2$	RR	BT	BH	BW	PS	E	V	M
31	M	15	106	63	87	99	16	36.7	165	50	6	4	5	6

The query subject (orange) was a 53-year-old female patient who suffered from abdominal pain in the upper-right quarter to right flanks for 3 days and noticed dizziness and tarry stool on the day of the interview. Through the retrieval, we obtained the 5 most similar patients (green) whose symptoms were hematuria, bloody stools, abdominal distention, abdominal pain, dizziness, and abdominal pain in the upper-right quarter.

In the subsequent finetuning based on the outcome, our proposed EDisease model demonstrated the highest performance among all compared models based on the area under the receiver operating characteristic of 0.876 (Table 2, Figure 4).

Table 2. Area under the receiver operating characteristic results.

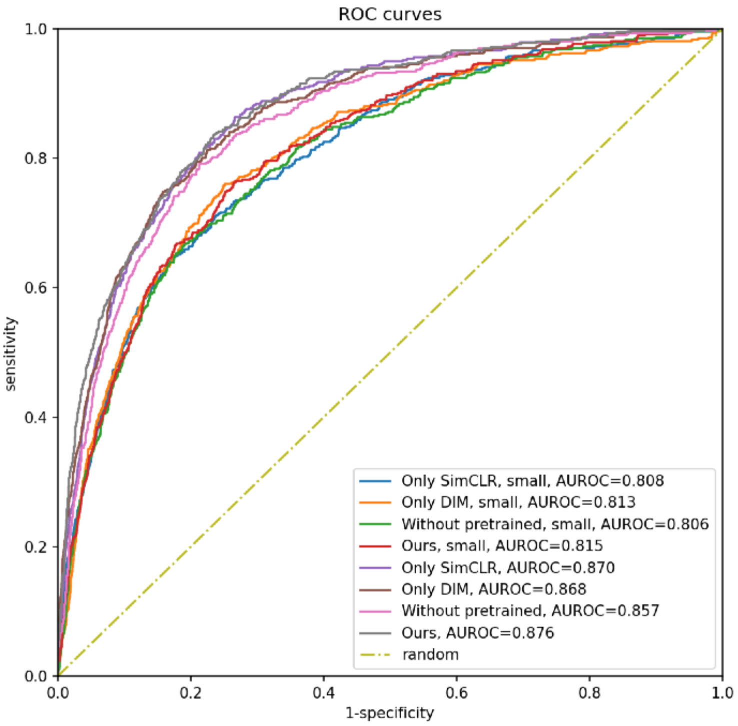
Model	Critical care outcome	
	NTUH-iMD[a]	NHAMCS[b]
Deep neural network [33] (age≤18)	N/A[c]	0.85
Deep neural network [32] (age≥18)	N/A	0.86
Hierarchical model [34]	N/A	0.84
Proposed model (without pretraining)	0.83	0.86
Proposed model	0.84	0.88

[a] NTUH-iMD: National Taiwan University Hospital Integrated Medical Database.

[b] NHAMCS: National Hospital and Ambulatory Medical Care Survey.

[c] N/A: not applicable.

Figure 4. Ablation study ROC curves. Both DIM and SimCLR can improve prediction performance through pretrained embedding. Small: denotes finetuned on the smaller data set. ROC: receiver operating characteristic. DIM: Deep InfoMax. SimCLR: Simple Contrastive Learning of Visual Representations. AUROC: area under the receiver operating characteristic.



The ablation study results demonstrate that our model outperformed those using fewer or no unsupervised methods, both on the large and small finetuning set. Unsurprisingly, the models finetuned on the smaller data set performed worse than those finetuned on the large National Hospital and Ambulatory Medical Care Survey training data set (Figure 4).

Discussion

The EDisease model performed well in the disease concept-embedding task and illustrated suitable clustering performance for the disease pattern. It also demonstrated promising results in the subsequent finetuning of the outcome.

The embedding of disease is meaningful. The enhanced label with the triage level indicates that the fifth level was clustered and illustrates some influence on different classification levels (see Multimedia Appendix 2). This embedding system can be used as a disease retrieval model, which encodes queries and finds the most relevant patients and diseases. In the retrieval demonstration, the query subject was a 53-year-old female patient who suffered from abdominal pain in the upper right quarter to right flanks for 3 days and noticed dizziness and tarry stool on the day of the interview. Through the retrieval, we obtained the five most similar patients with similar symptoms that were possibly related to different diseases (Figure 3). In another example of retrieval, the query subject was a 63-year-old male patient who suffered from sudden nausea, vomiting, and weakness and felt facial numbness at work. Patients with cardiogenic symptoms or neurologic symptoms were the retrieval results (see Multimedia Appendix 4). These results are meaningful for clinical practices, such as treating cardiogenic or neurological problems, as they characterize the most likely differential diagnoses.

Furthermore, because the multilingual BERT model is used as the text feature extractor, the proposed model can search for free text in English or Chinese. The retrieval results included relevant symptoms and possible diseases, likely from the co-occurrence relationship [27] of structural data or the feature extracted from the multilingual BERT model. The multilingual BERT model was pretrained again by the chief complaint and present illness pair. Even if we only use the chief complaint with structured data and medical history as input, the characteristics of the chief complaint may reflect those patients with the same complaint but different diseases discovered by the doctor.

The potential of the semantic vector on the EHR search engine has already been demonstrated [27]. Our main contribution to the development of EHR retrieval is the construction of patient-level conceptual embeddings. However, in some cases, such as complaints about epileptiform (query 12) and shingles-like skin lesions (query 6) (see Multimedia Appendix 3), retrieval performance was poor. They would also result from a co-occurrence relationship with structural data or medical records. In future work, more EHR data could be used to extend the model, which may enrich the performance for differential diagnosis and retrieval.

This pretrained model is also suitable for finetuning the outcome prediction. For the National Hospital and Ambulatory Medical Care Survey, our pretrained model outperformed the reference. The pretrained method of disease concept-embedding can improve the performance of subsequent tasks. The outcome prediction performance we obtained for National Hospital and Ambulatory Medical Care Survey data was slightly higher than for data from the National Taiwan University Hospital Integrated Medical Database. This result may be related to the structured chief complaints and comorbidities in the National Hospital and Ambulatory Medical Care Survey data. These structural features had been well extracted by humans using rule-based methods. In future work, we will use all medical histories instead of their average values or preprocess them using extractive summarization methods.

We proposed a model that uses limited information for embedding. The model is a useful contribution, given that the input data can be collected by people without a medical background. For instance, a sick person who has no idea what occurred to him or her can still collect information and feed it into the model to obtain more information about his or her problem. Simultaneously, the model can also help hospital staff determine a patient's disease with limited information. Most doctors (not only emergency department doctors) develop treatment plans based on limited information. Initially, doctors can obtain basic demographic information only after the patient has made an appointment. The doctor will then collect more information when the patient visits, including the chief complaint, symptoms, and medical histories. Furthermore, physical examination and some blood laboratory examinations, radiology, or ultrasound examinations are arranged to help establish a diagnosis and finalize the treatment plan. The disease concept was initially established by limited information and prompted doctors to arrange specific examinations; the disease concept will become more prominent. Similarly, disease embedding obtained through limited information can only be used for preliminary differential diagnosis. It can be further applied to suggest the most relevant information for the final diagnosis. After more information is appended to the embedding model, a precise diagnosis could be identified.

Some EHR models use ICD codes as inputs. However, problems may arise; for example, the disease might not have the correct ICD code in EHRs [9], or ICD coding rules might change, invalidating the model trained on the old ICD codes. In this study, we did not use ICD codes as inputs. In addition to the time-varying ICD coding rules, ICD coding requires experts, which is difficult for people with no medical background. Sometimes, if there is not enough information, even experienced doctors have difficulty with ICD coding. However, doctors can collect more information during treatment and observation, and then determine the disease and choose the most appropriate ICD code. Consequently, some of the ICD codes recorded in an EHR may be future information and may not be suitable for input to predictive models in some retrospective studies.

For machine learning methods used to embed disease concepts, vast amounts of patient data are required. Therefore, protecting personal privacy is a crucial issue in the research process. A leak of a patient's disease information would be disastrous and would infringe upon the patient's legal rights. Although several studies have focused on this issue, no accurate quantitative method exists to assess the privacy of research data sets [2]. For private EHR data sets that have not been evaluated for de-identification quality, encoding the patient's records and decoding them outside the hospital could create legal issues. In this study, DIM and SimCLR were used as the unsupervised methods, which have greater embedding capabilities [20] and do not train the decoder together, so there is no risk of leaking private information.

For the DIM method that maximizes mutual information, we assume each patient's visit is unique. This assumption is also critical for the SimCLR method, because the self-supervised learning methods require negative samples to maximize the embedding distance among groups. If the same input exists in the negative sample, it will disrupt the training process and cause failure to converge. This assumption is reasonable because everyone is undoubtedly unique. The only problem arises when the same patient revisits a health care facility because of the same disease. However, this problem can be solved by deleting duplicate accounts in the data set. For patients with several accounts, these visits can still be regarded as different due to different health conditions, chief complaints, or previous medical histories. We also assume that the number of diseases is limited and fewer than all visits. This assumption is also reasonable because the total number of visits is set to the size of the training set, and in medical experience, at least two patients suffer from the same disease.

The SimCLR method proposed a simple method to use a large amount of negative samples for self-supervised learning. In this study, ideally, we would have randomly masked the tokens in the input free text data. However, owing to graphic processing unit memory limitations and the heavy BERT model, if an ideal augmented method is adopted for a positive sample, only 8 samples could be used in each mini-batch, thereby sacrificing the advantage of a large number of negative samples. Consequently, we ended up adding a little noise to each m_i as a positive sample; in other words, we sacrificed the advantage of a positive sample.

Based on the ablation study results, both DIM and SimCLR can improve prediction performance through pretrained embedding. Although the improvement gap was small, the embedded pretraining method may be useful in future work. Furthermore, combining these two self-supervised methods can further improve prediction performance. These results are similar to those of DIM, which used two different discriminators on the training process and achieved a higher score [22].

Limitations

Although the National Taiwan University Hospital Integrated Medical Database included one medical center and two regional hospitals and had many patients and staff with extensive disease treatment experience, there is still a problem of “out-of-disease,” which indicates that the disease is not in the EHR data set. This problem will result in poor performance for newly diagnosed diseases (eg, COVID-19), but diseases with related symptoms can be found for reference. Because the National Taiwan University Hospital Integrated Medical Database data set was not public, we only recruited one doctor to evaluate the relevance of the search results. Although the evaluation results were subjective, no objective score exists (according to previous studies) for evaluating information retrieval performance. We recognize that other evaluation methods in this study might have been more meaningful.

Conclusions

The EDisease model uses limited information from patients and appropriately represents concept embedding. It can be further expanded as more data about the patient is collected.

The suitably-pretrained EHR model can be used as a medical experience retrieval system online in conjunction with the clinic staff. Moreover, it can be further finetuned to predict emergency events and enhance employees’ capabilities. The EDisease model could be widely adopted in the near future to help ease emergency department overcrowding: “ED-is-ease.”

Acknowledgments

We would like to thank the Ministry of Science and Technology, Taiwan, for financially supporting this research (grant MOST109–2634-F002–031, MOST109–2634-F002–041 and MOST 109–2634-F-002–029). We would also like to thank Huggingface for providing several excellent deep-learning codings. We are grateful to GitHub for providing the code repository used for the EDisease model.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Extended model.
[PNG File , 58 KB - jmir_v23i1e25113_app1.png]

Multimedia Appendix 2
Visualization of Embedding of Disease by t-SNE(2D). The embeddings were displayed in a list of (A) non-enhanced, (B) enhanced by gender, (C) age (10-year interval), (D) hospital level, (E) triage level, (F) whether a medical history exists, (G) outcome of ICU admission, and (H) death. There seems to be no specific difference in disease embedding based on (B) gender, (C) age> 10 years, (D) hospital grade, and (F) medical history. Triage level five, purple, was clustered together (E). In this disease embedding, the outcome of ICU or death did not appear to be linearly separable.
[PNG File , 4815 KB - jmir_v23i1e25113_app2.png]

Multimedia Appendix 3
Disease Retrieval Relevance. The relevance of each result to the query was judged by a seasoned doctor, who used the number of hits on the top five results to score from zero to five. Retrieval performance was high, but in some cases, such as shingles-like skin lesions (query 6) and epileptiform (query 12), the retrieval performance was poor.
[PNG File , 18 KB - jmir_v23i1e25113_app3.png]

Multimedia Appendix 4
Disease Retrieval Demonstration. The query subject (orange) was a 63-year-old male patient who suffered from sudden nausea, vomiting, and weakness and felt facial numbness at work. Through the retrieval, we obtained the top five similar patients (green)

whose symptoms were palpitation, chest pain, nausea, dizziness, anorexia, insomnia, and dizziness combined with limbs tremor. GCS: Glasgow coma scale. SBP: Systolic blood pressure. DBP: Diastolic blood pressure. HR: Heart rate. SpO_2: Blood oxygen saturation. RR: Respiratory rate. BT: Body temperature in Celsius. BH: Body height. BW: Body weight. PS: Pain scale. E: Eye response in Glasgow coma scale. V: Verbal response in Glasgow coma scale. M: Motor response in Glasgow coma scale. N/A : Not available.
[PNG File , 379 KB - jmir_v23i1e25113_app4.png]

References

1. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
2. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 01;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
3. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). Journal of Biomedical Informatics 2015 Jun;55:290-300. [doi: 10.1016/j.jbi.2015.05.003]
4. Chen Y, Chen Y, Lin J, Huang C, Lai F. Modified Bidirectional Encoder Representations From Transformers Extractive Summarization Model for Hospital Information Systems Based on Character-Level Tokens (AlphaBERT): Development and Performance Evaluation. JMIR Med Inform 2020 Apr 29;8(4):e17787. [doi: 10.2196/17787]
5. Thiessard F, Mougin F, Diallo G, Jouhet V, Cossin S, Garcelon N. RAVEL: retrieval and visualization in ELectronic health records. Netherlands: IOS Press BV; 2012:194-198.
6. Jin M, Li H, Schmid C, Wallace B. Using Electronic Medical Records and Physician Data to Improve Information Retrieval for Evidence-Based Care. In: Using Electronic Medical Records and Physician Data to Improve Information Retrieval for Evidence-Based Care. 2016 Presented at: 2016 IEEE International Conference on Healthcare Informatics (ICHI); Oct 4, 2016 - Oct 7, 2016; Chicago. [doi: 10.1109/ichi.2016.12]
7. International Statistical Classification of Diseases and Related Health Problems, 10th Revision. URL: https://icd.who.int/browse10/2016/en [accessed 2020-11-11]
8. Agarwal V, Podchiyska T, Banda JM, Goel V, Leung TI, Minty EP, et al. Learning statistical models of phenotypes using noisy labeled training data. J Am Med Inform Assoc 2016 May 12;23(6):1166-1173. [doi: 10.1093/jamia/ocw028]
9. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. BMJ 2015 Apr 24;350(apr24 11):h1885-h1885. [doi: 10.1136/bmj.h1885]
10. Li Y, Rao S, Solares JRA, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: Transformer for Electronic Health Records. Sci Rep 2020 Apr 28;10(1):online-online. [doi: 10.1038/s41598-020-62922-y] [Medline: 32346050]
11. Tran T, Nguyen TD, Phung D, Venkatesh S. Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). J Biomed Inform 2015 Apr;54:96-105 [FREE Full text] [doi: 10.1016/j.jbi.2015.01.012] [Medline: 25661261]
12. Beaulieu-Jones BK, Greene CS, Pooled Resource Open-Access ALS Clinical Trials Consortium. Semi-supervised learning of the electronic health record for phenotype stratification. J Biomed Inform 2016 Dec;64:168-178 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.007] [Medline: 27744022]
13. Wang Y, Wen A, Liu S, Hersh W, Bedrick S, Liu H. Test collections for electronic health record-based clinical information retrieval. JAMIA open 2019;2(3):360-368. [doi: 10.1093/jamiaopen/ooz016]
14. Mitra B, Craswell N. Neural Text Embeddings for Information Retrieval. 2017 Feb 23 Presented at: Neural text embeddings for information retrieval. Proceedings of the Tenth ACM International Conference on Web Search and Data Mining; February, 2017; Cambridge United Kingdom. [doi: 10.1145/3018661.3022755]
15. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K. Deep contextualized word representations. Deep contextualized word representations. Association for Computational Linguistics; New Orleans, Louisiana 2018:2227-2237. [doi: 10.18653/v1/n18-1202]
16. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Bert: Pre-training of deep bidirectional transformers for language understanding. Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019); 2019 Jun Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 2019; Minneapolis, Minnesota. [doi: 10.18653/v1/N19-1423]
17. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv. Preprint posted online July 28, 2020. URL: https://arxiv.org/abs/1910.10683 [accessed 2020-11-11]
18. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation 1997 Nov;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735]

19. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. ttention is all you need. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 57 Morehouse LaneRed HookNYUnited States: Curran Associates Inc; 2017:6000-6010.
20. D'Souza M, Van Munster CEP, Dorn JF, Dorier A, Kamm CP, Steinheimer S, et al. Autoencoder as a New Method for Maintaining Data Privacy While Analyzing Videos of Patients With Motor Dysfunction: Proof-of-Concept Study. J Med Internet Res 2020 May 8;22(5):e16669. [doi: 10.2196/16669]
21. Pattanayak S, Ludwig S. Improving Data Privacy Using Fuzzy Logic and Autoencoder Neural Network. In: Improving Data Privacy Using Fuzzy Logic and Autoencoder Neural Network. 2019 Presented at: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); June 23-26, 2019; New Orleans, LA, USA p. 1-6. [doi: 10.1109/fuzz-ieee.2019.8858823]
22. Hjelm R, Fedorov A, Lavoie-Marchildon S, Grewal K, Bachman P, Trischler A. Learning deep representations by mutual information estimation and maximization. 2019 Presented at: International Conference for Learning Representations (ICLR), 2019; 2019; New Orleans, Louisiana URL: https://arxiv.org/abs/1808.06670
23. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. 2020 Oct Presented at: Proceedings of the 28th ACM International Conference on Multimedia; Seattle, WA. USA2020; October, 2020; New YorkNYUnited States p. 2193-2210. [doi: 10.1145/3394171.3413694]
24. Salton G, Wong A, Yang CS. A vector space model for automatic indexing. Commun. ACM 1975 Nov;18(11):613-620 [FREE Full text] [doi: 10.1145/361219.361220]
25. Ye C, Fabbri D. Extracting similar terms from multiple EMR-based semantic embeddings to support chart reviews. Journal of Biomedical Informatics 2018 Jul;83:63-72. [doi: 10.1016/j.jbi.2018.05.014]
26. Biron P, Metzger M, Pezet C, Sebban C, Barthuet E, Durand T. An information retrieval system for computerized patient records in the context of a daily hospital practice: the example of the Léon Bérard Cancer Center (France). Appl Clin Inform 2017 Dec 20;05(01):191-205. [doi: 10.4338/aci-2013-08-cr-0065]
27. Hanauer DA, Wu DT, Yang L, Mei Q, Murkowski-Steffy KB, Vydiswaran VV, et al. Development and empirical user-centered evaluation of semantically-based query recommendation for an electronic health record search engine. Journal of Biomedical Informatics 2017 Mar;67:1-10. [doi: 10.1016/j.jbi.2017.01.013]
28. Maaten L, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. (Nov) 2008;9:2579-2605 [FREE Full text]
29. Zheng J, Yu H. Methods for linking EHR notes to education materials. Inf Retrieval J 2015 Sep 3;19(1-2):174-188. [doi: 10.1007/s10791-015-9263-1]
30. Sklar DP, Crandall CS, Loeliger E, Edmunds K, Paul I, Helitzer DL. Unanticipated death after discharge home from the emergency department. Ann Emerg Med 2007 Jun;49(6):735-745. [doi: 10.1016/j.annemergmed.2006.11.018] [Medline: 17210204]
31. Obermeyer Z, Cohn B, Wilson M, Jena AB, Cutler DM. Early death after discharge from emergency departments: analysis of national US insurance claims data. BMJ 2017 Feb 01:j239. [doi: 10.1136/bmj.j239]
32. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019 Feb 22;23(1):open-access. [doi: 10.1186/s13054-019-2351-7]
33. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine Learning–Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. JAMA Netw Open 2019 Jan 11;2(1):e186937. [doi: 10.1001/jamanetworkopen.2018.6937]
34. Barnes S, Saria S, Levin S. An Evolutionary Computation Approach for Optimizing Multilevel Data to Predict Patient Outcomes. Journal of Healthcare Engineering 2018;2018:1-10. [doi: 10.1155/2018/7174803]
35. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JMC. Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: a Review. Artif Intell Med 2020 Jan;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
36. National Center for Health Statistics. Ambulatory Health Care Data. Centers for Disease Control and Prevention. 2021 Jan 05. URL: https://www.cdc.gov/nchs/ahcd/index.htm [accessed 2021-09-26]
37. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A. Transformers: State-of-the-Art Natural Language Processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations 2020 Oct:38-45. [doi: 10.18653/v1/2020.emnlp-demos.6]
38. Vincent P, Larochelle H, Bengio Y. Extracting and composing robust features with denoising autoencoders. In: ICML '08: Proceedings of the 25th international conference on Machine learning. 2008 Jul Presented at: Proceedings of the 25th international conference on Machine learning; July 2008; Helsinki, Finland p. 1096-1103. [doi: 10.1145/1390156.1390294]
39. Gutmann M, Hyvärinen A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings. 2010 May 13 Presented at: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; 13-15 May 2010; Sardinia, Italy p. 297-304 URL: https://www.jmlr.org/papers/volume13/gutmann12a/gutmann12a.pdf
40. Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding. arXiv preprint arXiv. 2018. URL: https://arxiv.org/abs/1807.03748 [accessed 2020-09-15]

41. Nowozin S, Cseke B, Tomioka R. f-gan: Training generative neural samplers using variational divergence minimization. In: NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016 Dec 06 Presented at: NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems; Dec 4, 2016 - Dec 9, 2016; Barcelona, SPAIN p. 271-279. [doi: 10.5555/3157096.3157127]
42. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Communications of the ACM 2020 Oct 22;63(11):139-144. [doi: 10.1145/3422622]
43. Kingma D. Adam: A Method for Stochastic Optimization. In: 3rd International Conference on Learning Representations, ICLR 2015. 2015 May 07 Presented at: 3rd International Conference on Learning Representations, ICLR 2015; May 7-9, 2015; San Diego, CA, USA.
44. EDisease_NHAMCS. URL: https://github.com/wicebing/EDisease_NHAMCS [accessed 2021-11-11]

Abbreviations

AUROCs: Area Under the Receiver Operating Characteristics
BERT: Bidirectional Encoder Representations from Transformers
DIM: Deep InfoMax
EHRs: electronic health records
ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision
SimCLR: Simple Contrastive Learning of Visual Representations
t-SNE: t-distributed stochastic neighbor embedding

Edited by G Eysenbach; submitted 18.10.20; peer-reviewed by N Gronewold, E Choi; comments to author 06.11.20; revised version received 19.11.20; accepted 15.01.21; published 27.01.21.

Please cite as:
Chen YP, Lo YH, Lai F, Huang CH
Disease Concept-Embedding Based on the Self-Supervised Method for Medical Information Extraction from Electronic Health Records and Disease Retrieval: Algorithm Development and Validation Study
J Med Internet Res 2021;23(1):e25113
URL: http://www.jmir.org/2021/1/e25113/
doi:10.2196/25113
PMID:33502324

Original Paper

A Newly Developed Web-Based Resource on Genetic Eye Disorders for Users With Visual Impairment (Gene.Vision): Usability Study

Jian Lee Yeong[1*], MBChB; Peter Thomas[2*], MA, BM, BChir, FRCOphth, PhD; James Buller[3], BSc; Mariya Moosajee[1,2,4], BSc, MBBS, PhD, FRCOphth

[1]Institute of Ophthalmology, University College London, London, United Kingdom
[2]Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom
[3]Aniridia Network, Sheffield, United Kingdom
[4]Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Mariya Moosajee, BSc, MBBS, PhD, FRCOphth
Institute of Ophthalmology
University College London
11-43 Bath Street
London, EC1V 9EL
United Kingdom
Phone: 44 2076086800
Email: m.moosajee@ucl.ac.uk

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2021/1/e27330/

Abstract

Background: Despite the introduction of the Web Content Accessibility Guidelines and legislations, many websites remain poorly accessible to users with disability, especially those with visual impairment, as the internet has become a more visually complex environment. With increasing reliance on the internet and almost 2 million people in the United Kingdom being affected by vision loss, it is important that they are not overlooked when developing web-based materials. A significant proportion of those affected have irreversible vision loss due to rare genetic eye disorders, and many of them use the internet as a primary source of information for their conditions. However, access to high-quality web-based health information with an inclusive design remains a challenge for many. We have developed a new web-based resource for genetic eye disorders called Gene.Vision that aims to provide a holistic guide for patients, relatives, and health care professionals.

Objective: Through a usability testing session of our website prototype, this study aims to identify key web-based accessibility features for internet users with vision impairment and to explore whether the contents provided in Gene.Vision are relevant and comprehensible.

Methods: A face-to-face testing session with 8 participants (5 patients, 2 family members, and 1 member of the public) and 8 facilitators was conducted on a prototype website. Remote testing was performed with another patient due to COVID-19 restrictions. Home page design, navigation, content layout and quality, language, and readability were explored through direct observation and task completion using the think-aloud method. A patient focus group was organized to elicit further feedback. Qualitative data were gathered and analyzed to identify core themes through open and axial coding.

Results: All participants had good computer literacy; 6 patients with visual impairment used visual aid software including iOS VoiceOver and Speak Screen, iOS Classic Invert, ZoomText 2020, Job Access With Speech, and Nonvisual Desktop Access. The features identified by the participants that will enhance accessibility and usability for users with visual impairment were a consistent website layout, a structured information hierarchy with a clear description of links, good chromatic and luminance contrast, a simple home page with predictable and easy navigation, adaptability to various assistive software, and readable and relevant content. They reported that dynamic content (such as carousels) and large empty spaces reduced accessibility. Information on research, support available, practical advice, and links to charities were incentives for repeated website visits.

Conclusions: We demonstrated the importance of developing a website with a user-based approach. Through end user testing, we identified several key web-based accessibility features for people with visual impairment. Target end users should always be involved early and throughout the design process to ensure their needs are met. Many of these steps can be implemented easily and will aid in search engine optimization.

(J Med Internet Res 2021;23(1):e19151) doi:10.2196/19151

KEYWORDS

internet access; blindness; eye disease; genetic diseases; usability testing; qualitative research; internet-based intervention; consumer health information; mobile phone

Introduction

Background

People with disabilities have been recognized to experience inequalities in multiple aspects of their lives from education and employment to health care, finances, and leisure [1], with the internet being a key accessibility feature in all areas. It was intended to be user-friendly for every individual regardless of any mental or physical disability but poor design and coding of what is usually thought of as a graphical user interface poses difficulties, particularly for users with visual impairment [2-4]. With an increasing reliance on the internet and almost 2 million people in the United Kingdom being affected by vision loss, of which 360,000 are registered with visual impairment or severe visual impairment (blind), it is important that they are not overlooked when developing web-based materials lest they risk being socially excluded altogether [5,6].

To make the internet more accessible, the World Wide Web Consortium first introduced the Web Content Accessibility Guidelines (WCAG) in 1999; its current iteration (WCAG 2.1), published in 2018, serves as an update rather than as a replacement of WCAG 2.0 (introduced in 2008). It contains 13 guidelines based on 4 main principles: perceivable, operable, understandable, and robust, with each guideline having 3 levels of testable success criteria—A (lowest level of accessibility) and AA and AAA (highest level of accessibility) [7]. Both the United Kingdom and the European Union have adopted it into their legislation, making it a legal requirement for public sector websites to be accessible to people with disabilities [8,9].

Despite this, only a small number of websites are compliant with the UK accessibility standard of WCAG 2.1 AA [10]. With consumers increasingly turning to the internet as a primary source of health-related information, websites should not only contain relevant and helpful information for patients but should also not exclude anyone with a disability. To help developers of web-based health platforms achieve this, the UK National Health Service (NHS) published a Digital Service Manual focusing on building consistent, usable, and accessible services using a patient-centric approach [11]. It provides guidance on various aspects of web design, such as conducting user research, building user interface prototypes, content styles, and optimizing accessibility.

A significant group of web-based health information consumers are those affected by rare diseases that together affect 1 in 17 people [12]. Owing to the nature of their conditions, their physicians often lack the experience or familiarity to provide expert care for them. They end up searching the web to keep abreast of current research and look for support through online communities. However, the quality of these websites tends to be poor, either they lack credibility or the contents are not kept up-to-date or pitched at the right level [13]. Furthermore, most of these sites are found to be poorly accessible to users with visual impairment, which can have a profound impact on patients affected by rare genetic eye disorders who usually have visual disability [13,14]. Inherited retinal diseases are the most common causes of severe visual impairment registration among working adults in England and Wales, whereas globally, 60% of blindness among infants is due to genetic diseases [15,16].

Objectives

To fill this void, we are developing a new web-based resource for these conditions with information written in 2 formats: one using more lay terms for patients and families, while the other has more technical details tailored toward health care professionals. It aims to provide a holistic resource, including the causative gene or genes of a condition, current research, management, and support services. Most importantly, we prioritized creating a usable and highly accessible website for our target end users, who mainly have visual impairment.

Although automated testing tools can help identify some of the website's accessibility barriers, most of the WCAG 2.1 guidelines require human testing and judgment. Failing to trial a specific design with users with disability will result in a website with limited *usable accessibility*, as the human interface aspect is not evaluated [17]. Therefore, the NHS Digital Service Manual recommends that research and testing with real-world users are performed early and throughout the design process [18]. In line with this recommendation, we conducted a usability testing session on a prototype of our website with patients to identify website features that would increase accessibility for users with visual impairment. Our secondary objective is to explore whether the contents provided on our website were relevant and comprehensible to both patients and clinicians.

Methods

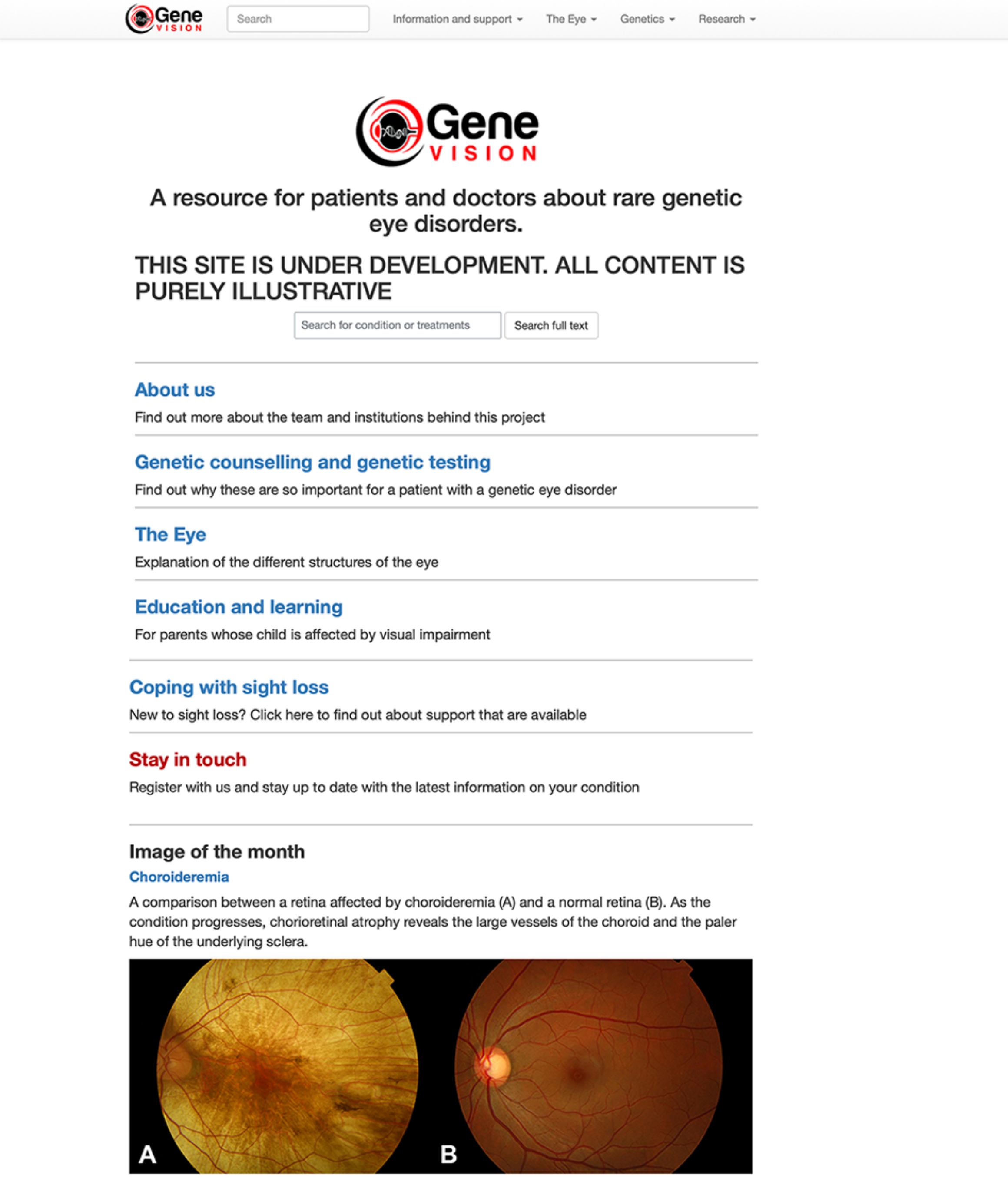
Prototype

Before development, we spoke to patients with genetic eye disorders and their families to identify potential *pain points*, which were websites with poor accessibility, unreliable and outdated information, contents written with excess medical jargon and insufficient material on research, practical advice, and available support. An accessibility specialist, who was a patient herself (see the Acknowledgments section), was also consulted to identify key accessibility features for those with visual impairment. On the basis of feedback from these interactions, we built a prototype of our website using the Wikimedia platform (Wikimedia Foundation, Inc) [19], which contained the following components.

Home Page

The home page's navigation bar is organized into 4 main items: information and support, the eye, genetics, and research (Figure 1). The subcategories of each main item are listed in Textbox 1 along with a description for each category. There is a search box in the center whereupon a search term or terms is entered, all pages containing the term or terms are displayed on a new page (Figure 2).

Figure 1. Home page of the Gene.Vision Wikimedia prototype. The navigation bar serves as the header of the page. The search box is above the listed menu items. Our participants pointed out that the navigation bar and search box had poor chromatic and luminance contrast. The large area of white space on the right in portrait mode can disorientate magnifier users.



A resource for patients and doctors about rare genetic eye disorders.

THIS SITE IS UNDER DEVELOPMENT. ALL CONTENT IS PURELY ILLUSTRATIVE

About us

Find out more about the team and institutions behind this project

Genetic counselling and genetic testing

Find out why these are so important for a patient with a genetic eye disorder

The Eye

Explanation of the different structures of the eye

Education and learning

For parents whose child is affected by visual impairment

Coping with sight loss

New to sight loss? Click here to find out about support that are available

Stay in touch

Register with us and stay up to date with the latest information on your condition

Image of the month

Choroideremia

A comparison between a retina affected by choroideremia (A) and a normal retina (B). As the condition progresses, chorioretinal atrophy reveals the large vessels of the choroid and the paler hue of the underlying sclera.

Textbox 1. Subcategories of each main item on the navigation bar in the Gene.Vision prototype with a brief explanation of each category.

Information and support

- Coping with sight loss: outlining support services that are available for patients recently diagnosed with vision loss. It also explains the roles of low vision clinic and eye clinic liaison officers and directs patients to resources on assistive technology
- Registration for sight impairment: the process of certification and registration are outlined here along with the associated advantages
- Driving: the minimum driving standards in the United Kingdom are explained here with further links to the government information portal. Patients are also signposted to alternative ways of transportation if they are no longer able to drive
- Education and learning: the various supports that are available for children with visual impairment are listed here with advice on how to access these services
- Charles Bonnet Syndrome (CBS): an explanation about CBS, coping mechanisms, current research and links to organizations offering support to those affected

The Eye (pages providing lay explanations of the following ocular structures):

- Conjunctiva
- Sclera
- Cornea
- Trabecular meshwork
- Iris and pupil
- Lens
- Vitreous humor
- Retina
- Optic nerve
- Choroid

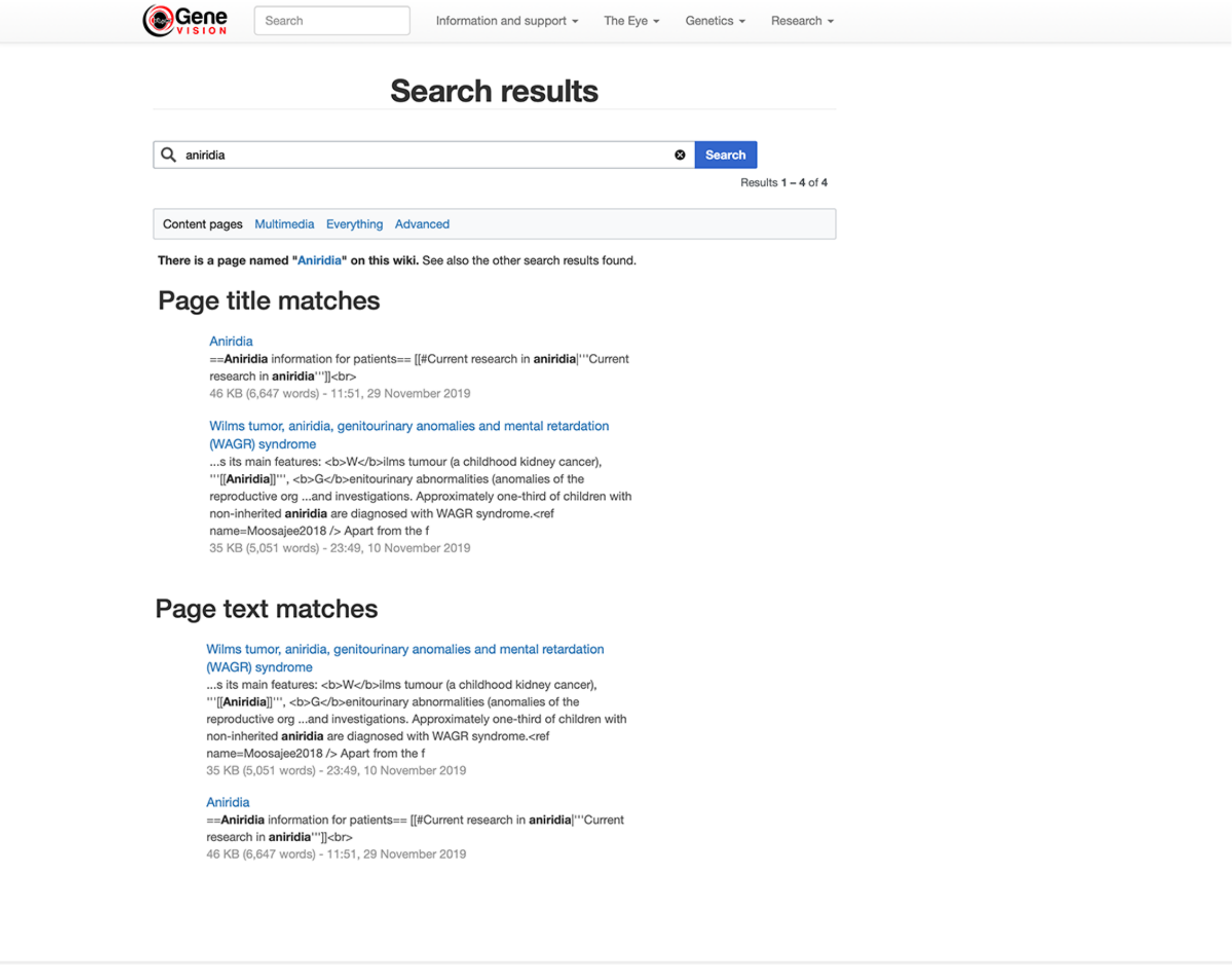
Genetics

- Introduction to genetics: a brief overview of DNA, genes, and chromosomes
- Inheritance patterns: explanations on the more common inheritance patterns encountered in genetic eye disorders (autosomal dominant, autosomal recessive, X-linked, and mitochondrial)
- Genetic counseling and genetic testing: the role of genetic counselors, the process of genetic testing and its associated benefits and limitations are explained in this page

Research

- Clinical trials: explaining the different aspects of clinical trials including the various phases, the role of ethics committees, and the benefits of participation
- Treatments under research: this section contains various investigative therapies for genetic eye disorders with links to their corresponding pages. Only gene therapy was prepared in this prototype
- Participate in research: This link aims to connect to a website listing all the current ongoing trials in Moorfields Eye Hospital. This website has not been set up yet during the testing session nor at the time of writing this manuscript

Figure 2. Search results for the term aniridia. Our participants found the display confusing compared to what they are used to from popular search engines.



Search results

Results 1 – 4 of 4

There is a page named "Aniridia" on this wiki. See also the other search results found.

Page title matches

Aniridia
==**Aniridia** information for patients== [[#Current research in **aniridia**|'''Current research in **aniridia**''']]
46 KB (6,647 words) - 11:51, 29 November 2019

Wilms tumor, aniridia, genitourinary anomalies and mental retardation (WAGR) syndrome
...s its main features: Wilms tumour (a childhood kidney cancer), '''[[**Aniridia**]]''', Genitourinary abnormalities (anomalies of the reproductive org ...and investigations. Approximately one-third of children with non-inherited **aniridia** are diagnosed with WAGR syndrome.<ref name=Moosajee2018 /> Apart from the f
35 KB (5,051 words) - 23:49, 10 November 2019

Page text matches

Wilms tumor, aniridia, genitourinary anomalies and mental retardation (WAGR) syndrome
...s its main features: Wilms tumour (a childhood kidney cancer), '''[[**Aniridia**]]''', Genitourinary abnormalities (anomalies of the reproductive org ...and investigations. Approximately one-third of children with non-inherited **aniridia** are diagnosed with WAGR syndrome.<ref name=Moosajee2018 /> Apart from the f
35 KB (5,051 words) - 23:49, 10 November 2019

Aniridia
==**Aniridia** information for patients== [[#Current research in **aniridia**|'''Current research in **aniridia**''']]
46 KB (6,647 words) - 11:51, 29 November 2019

To highlight certain relevant topics to users, some of the subcategories from the main navigation bar were placed into a list below the search box, each having a brief excerpt (Figure 1).

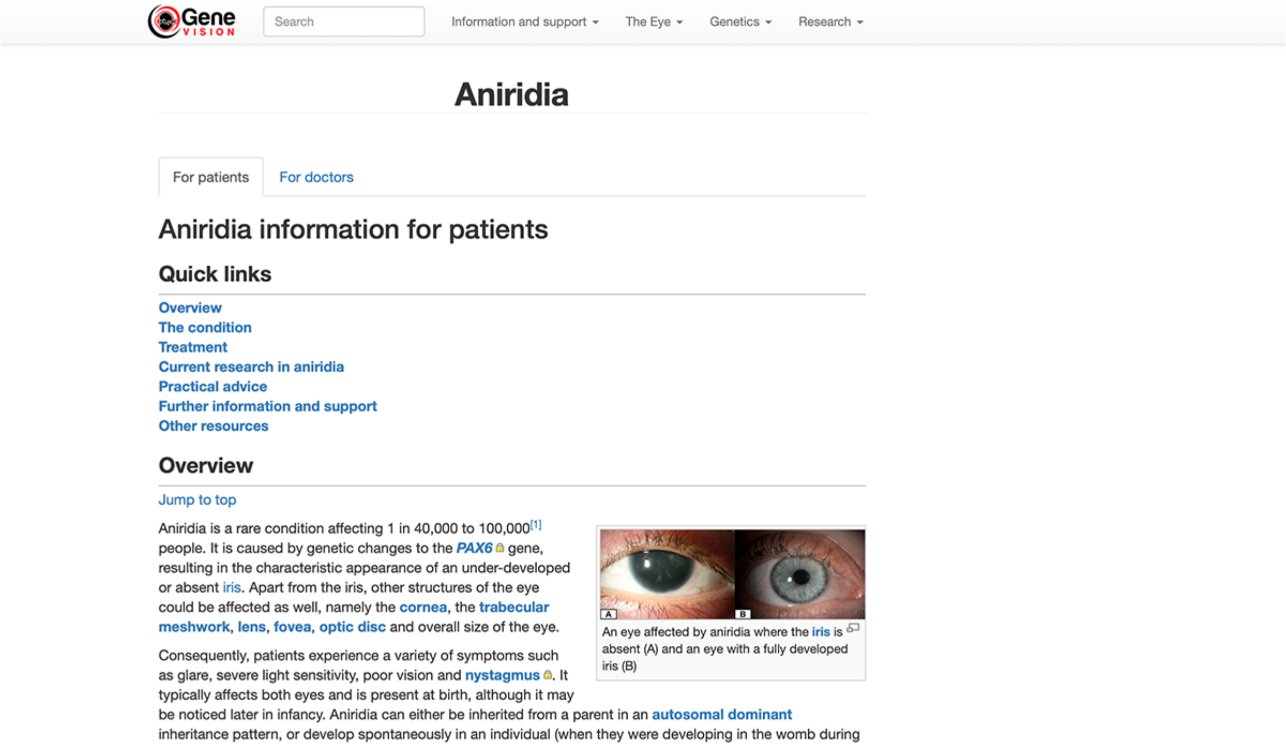
In addition to the Wikimedia-based home page, we also designed another home page prototype using WordPress for the purpose of this testing. The screen recordings of both home page designs are available in Multimedia Appendices 1 and 2.

Content Pages

Conditions

Two versions are available for each condition page: one for the patients and general public, and the other for health care professionals. Users can toggle between these 2 versions using the tabs below the main page title (Figure 3).

Figure 3. Example of a condition page. Users can toggle between the patient and health care professional versions using the For patients and For doctors tabs. A table of contents is placed under the Quick links section, where users can go directly to the topic of interest without additional scrolling. Jump-to-top links are available under every heading to facilitate quick navigation back to the Quick links section.



Aniridia

Aniridia information for patients

Quick links

Overview
The condition
Treatment
Current research in aniridia
Practical advice
Further information and support
Other resources

Overview

Jump to top

Aniridia is a rare condition affecting 1 in 40,000 to 100,000[1] people. It is caused by genetic changes to the *PAX6* gene, resulting in the characteristic appearance of an under-developed or absent iris. Apart from the iris, other structures of the eye could be affected as well, namely the cornea, the trabecular meshwork, lens, fovea, optic disc and overall size of the eye.

An eye affected by aniridia where the iris is absent (A) and an eye with a fully developed iris (B)

Consequently, patients experience a variety of symptoms such as glare, severe light sensitivity, poor vision and nystagmus. It typically affects both eyes and is present at birth, although it may be noticed later in infancy. Aniridia can either be inherited from a parent in an autosomal dominant inheritance pattern, or develop spontaneously in an individual (when they were developing in the womb during

The former focuses on aspects that we believe patients will find helpful, such as current research, information on available support, and links to both umbrella and condition-specific charities. When possible, we asked a patient representative from a charity related to the condition to review our material for readability. The health care professional version focuses more on the clinical presentation, investigations, general management guidance including genetic counseling, genetic testing, and updates on current research. Full citations are listed in both versions.

A total of 3 conditions were prepared for this prototype: choroideremia; aniridia; and Wilms tumor, aniridia, genitourinary anomalies, and range of developmental delay (WAGR) syndrome. The patient versions of the choroideremia and aniridia pages were reviewed by patient representatives from Retina UK and Aniridia Network, respectively. The review of the WAGR syndrome page was still pending at the time of testing.

Other Pages

For pages listed in Textbox 1, only the patient or general public version was available in the prototype.

The pages that were prepared for this prototype were as follows: *Coping with sight loss*, *Driving*, all the pages listed under *The eye*, *Genetic counseling and genetic testing*, *Clinical trials*, and *Gene therapy*.

To improve user navigation, a table of contents is listed at the top of each page termed *Quick links* (Figure 3). Users are able to navigate to each subheading by clicking on the topic of interest. They can easily navigate back to *Quick links*, by using the *Jump to top* link. A screen recording of this function is available in Multimedia Appendix 3. Each page always starts with a brief overview of the selected topic, with the structure and font used consistently throughout the website. The information provided on all pages is based on current UK practices. All images on the website were accompanied with the alt-text function, except those that were for decorative purposes only. The prototype was tested manually with the NonVisual Desktop Access screen reader (NV Access) by one of the authors (JY) before the usability testing session to ensure compatibility and troubleshoot any fundamental accessibility barriers (eg, unclear alt-text, missing or ambiguous links).

Usability Testing

Participants were recruited through the vision loss charity Retina UK and from the senior author's (MM) genetic eye disease clinic in Moorfields Eye Hospital NHS Foundation Trust, London.

A face-to-face usability testing session and patient focus group (1 participant to 1 facilitator) of our prototype was conducted at the University College of London Institute of Ophthalmology computer laboratory on November 11, 2019. Participants were encouraged to bring their personal devices that they usually used to access the web. For those who chose to use the laboratory's Windows PC desktops, assistive software was installed upon prior request. They also chose which web browser to use based on their experience and comfort level. The device, web browser, and assistive software used to access the website were recorded by the facilitators.

A short survey (Multimedia Appendix 4) was conducted at the beginning of the session to gather information about the demographics and computer literacy of each participant. This was assessed using a 5-point Likert scale based on the UK government digital inclusion scale [20]:

- Not at all confident: someone who does not know how to use the internet at all.
- Slightly confident: able to perform tasks on the internet with full guidance.
- Somewhat confident: able to perform certain tasks independently, does not tend to deviate from these tasks when using the internet (task-specific).
- Very confident: someone considered to have basic digital skills and able to use the internet effectively.
- Extremely confident: someone who usually has a background in computing or is able to code or design content.

The last 2 questions of the survey focused on the participant's current source of information about their condition and the type of information they considered important but currently lacking.

During the testing session, we used 2 usability techniques: direct observation and task completion. The tasks outlined for the participants (Textbox 2) were designed to identify specific features and issues pertaining to the home page design, site navigation, and content. For our condition pages targeted to the patients and general public, we used identical pages from Wikipedia as a benchmark because it is one of the most viewed medical resources globally and tends to appear as one of the top results in general search engines for keywords associated with rare diseases compared with other web-based health resources [21,22]. Furthermore, Wikipedia has a standardized information structure optimized for accessibility (WACG 2.0) and contributors are encouraged to follow the guidance provided in their *Manual of Style* to maintain this standard [23].

Textbox 2. The tasks set out for the participants of the usability testing session.

Home page

1. Home page design: Participants were asked to compare the 2 home page designs, Wikimedia (Wikimedia Foundation, Inc) and WordPress, and highlight the aspects that improved or affected accessibility for each design
2. Navigation: navigating to the *Driving, Retina,* and *Genetic counselling* and *genetic testing* pages by using a combination of the navigation bar, search box, and the listed home page menu. While testing the navigation bar, we wanted to find out if participants preferred a drop-down menu displaying the full list of items (eg, *Information and Support* section) or a tab that links to a separate page containing the full list of items (eg, *The Eye* section). Participants were asked to explore each of the assigned pages and assess the relevance and readability of the content provided

Condition pages

- Participants were required to read 2 condition pages on our website: choroideremia and aniridia. They were asked to compare identical pages from Wikipedia with regards to information layout, readability, and quality of content. They then had to point out features that they liked and disliked from both the prototype and Wikipedia

While performing the assigned tasks, participants were encouraged to verbalize their initial impressions of the website, its accessibility and ease of use, comprehensibility of the content, and suggestions for improvement (think-aloud method). They were prompted by the facilitators to explain their actions and expectations. All responses were recorded by each facilitator in a standardized proforma. The facilitators primarily served as passive observers (direct observation) and were only allowed to assist if their respective participant reached an obstacle, at the same time noting how the participant encountered this.

A patient focus group for all the participants was conducted by the main facilitator at the end of the usability testing session to encourage feedback and discussion. Participants were asked about their general thoughts about the website, whether the current version was usable by individuals with visual impairment, any features that they liked or disliked, and if there were any changes or additions they would like to see. Further discussions were prompted by asking participants to expand their answers and seek opinions or counter-opinions from the rest of the group. The focus group was audio taped and transcribed after obtaining informed consent from all participants. A summary of the discussions was prepared after recording by the main author (JY).

Clinician Testing

The website was also created as a resource for health care professionals who may see patients with genetic eye diseases and require further information on the condition and management advice. Hence, the content was assessed remotely by a range of health care professionals from various disciplines, including a consultant general practitioner, a consultant pediatrician, 2 ophthalmology trainees, 3 ophthalmology consultants of different subspecialties (general, pediatrics, and medical retina), and a genetic counselor. Their assigned tasks were similar to those outlined in Textbox 2, with particular focus on home page navigation, information layout, and content quality in the professional versions of choroideremia, aniridia, and WAGR syndrome. Feedback was recorded in the same proforma used in the usability testing session.

Qualitative Data Analysis

The data collected in the standardized proformas and the audio transcripts from the patient focus group were analyzed by one author (JY) in an inductive manner rather than theory based, as we wanted to understand participants' experiences with the prototype. Codes were attributed to various key phrases and paragraphs in both the written and audio transcripts (open coding) to identify core themes. Axial coding was then applied to develop connections between the open codes. Coding was performed with the qualitative software program NVivo 12 (QSR International).

Results

Participant Characteristics

There were 8 participants in our usability testing session and their demographics are presented in Table 1. A participant (P9), who used the popular Job Access With Speech (JAWS) screen reader, tested the website remotely during the COVID-19 lockdown period in the United Kingdom. In total, 6 participants were affected by genetic eye disorders, while 2 were parents of a patient with type II Usher syndrome. One participant had full sight and worked as an information specialist in a library. Most participants had basic digital skills, but 3 participants were considered experts. One participant considered herself a *task-specific* web user. The devices, web browsers, and assistive software used by the participants are shown in Table 2. The full results of the pretesting survey (including P9) are presented in Multimedia Appendix 5.

Table 1. Demographics of the participants in the usability testing session.

Participant	Gender	Age (years)	Employment status	Internet usage	Confidence in using the internet	Health status	Diagnosis
P1	M[a]	53	Employed	Yes	Very confident	Unaffected parent	Son-type II Usher syndrome
P2	F[b]	52	Employed	Yes	Very confident	Unaffected parent	Son-type II Usher syndrome
P3	F	41	Unemployed	Yes	Somewhat confident	Patient	Retinitis pigmentosa
P4	F	62	Retired	Yes	Very confident	Patient	Cone-rod dystrophy
P5	M	40	Employed	Yes	Extremely confident	Patient	Aniridia
P6	M	37	Employed	Yes	Very confident	Patient	Retinitis pigmentosa
P7	F	25	Employed	Yes	Extremely confident	Patient	Type II Usher syndrome
P8	M	45	Employed	Yes	Extremely confident	Unaffected individual	N/A[c]
P9[d]	M	70	Retired	Yes	Very confident	Patient	Cone-rod dystrophy

[a] M: Male.

[b] F: Female.

[c] N/A: not applicable.

[d] P9 was only able to test the website remotely due to COVID-19 lockdown rules.

Table 2. Device, web browser, and assistive software used by each participant.

Participant	Device	Web browser	Assistive software
P1	Windows PC	Google Chrome (Google LLC)	N/A[a]
P2	Windows PC	Google Chrome (Google LLC)	N/A
P3	iPhone	Safari (Apple Inc)	iOS VoiceOver (Apple Inc)
P4	Windows PC	Mozilla Firefox (Mozilla Foundation)	Zoomtext 2020 (Freedom Scientific) with invert color setting
P5	Windows PC	Google Chrome (Google LLC)	Windows magnifier at 200% magnification (Microsoft)
P6	iPad	Safari (Apple Inc)	iOS Classic Invert, iOS VoiceOver (Apple Inc)
P7	iPad	Safari (Apple Inc)	iOS Large Text, color filters, iOS VoiceOver and Speak Screen (Apple Inc)
P7	Windows PC	Google Chrome (Google LLC)	NVDA[b]
P8	Windows PC	Google Chrome (Google LLC)	N/A
P9	Windows PC	Google Chrome (Google LLC)	JAWS[c]

[a] N/A: not applicable.

[b] NVDA: NonVisual Desktop Access.

[c] JAWS: Job Access With Speech.

Qualitative Analysis

Participant Feedback

The positive features of the website that enhanced accessibility for users with visual impairment are outlined in Textbox 3. One of the main features highlighted by the participants was the consistent layout throughout the website. These included having left-aligned texts, using a consistent font and color along with having the Quick links section in all pages. They identified that having a zigzag alignment of text, as in the case of the WordPress-designed home page (Multimedia Appendix 2), limited accessibility. For the content pages, our participants noted that having clear and concise information helped with accessibility. These included having accurate and logically ordered headings, arranging information using bullet points for ease of reading (Figure 4) and the purpose of each link clearly identified from the link text alone. Another important aspect that favored accessibility was good adaptability to various assistive software and devices, especially in this day and age where mobile devices are increasingly popular and accessible. The participants also favored a simple home page that was easy to navigate to other pages on the website. Participants with and without visual impairment liked that minimal scrolling is required on the home page. They also particularly liked the *Quick links* and *Jump to top* functions, as they offered straightforward navigation. In addition, they stated that content written in an easily understandable language and arranged in a structured manner added value to the website and improved the overall user experience.

Textbox 3. Positive accessibility features of the website.

Consistent layout and navigation:

"I like the uniform layout and colour theme. The text all have the same size throughout which is good as I did not have to change my magnification level often. The left-aligned texts make it easier to read as well. The texts on the Wordpress design can be easily missed on magnified view as they are not all properly aligned." [P4, 62-year-old cone-rod dystrophy patient]

"The zig-zag arrangement of icons in the Wordpress design can be bad for accessibility." [P5, 40-year-old patient with aniridia]

Structured information hierarchy with a clear description of links:

"The headings and subheadings of each page are very clear. Each heading describes exactly what is written." [P8, 45-year-old unaffected individual]

"I found it easy to navigate through the different headings for each page using the 'Insert-F6' function on JAWS which lists all the headings." [P9, 70-year-old patient with cone-rod dystrophy]

"I like the bullet point arrangement which made it easy to read. The targets of the links are also very clear with accurate descriptions." [P5, 40-year-old patient with aniridia]

"The Wikipedia articles have too long paragraphs; having bullet points will make it easier to read." [P1, 53-year-old unaffected father of a patient with type II Usher syndrome]

"The links are relevant in gene.vision and work well. There are too many links to read with a screen reader on Wikipedia." [P6, 37-year-old patient with retinitis pigmentosa]

Adaptability to different assistive software:

"The website works really well and is really fluid when using Voiceover on my iPhone." [P3, 41-year-old patient using iOS Voiceover with retinitis pigmentosa]

"I am still able to see the whole page at 200% magnification." [P5, 40-year-old patient using Windows magnifier with aniridia]

"The website works well with my Zoomtext settings." [P4, 62-year-old patient using ZoomText 2020 with cone-rod dystrophy patient]

"I found the website easy to understand and JAWS friendly." [P9, 70-year-old patient using JAWS with cone-rod dystrophy]

Simple home page with easy navigation:

"The home page is all on one page, which I like. No sliding or scrolling required." [P1, 53-year-old unaffected father of a patient with type II Usher syndrome]

"The drop-down menu at the top works well and as expected. The simpler the navigation the better. Gene.vision is easier to navigate through compared to Wikipedia." [P3, 41-year-old patient with retinitis pigmentosa]

"I like the drop-down menu as less scrolling is involved. I really like the 'Jump to top' button to navigate around when reading." [P4, 62-year-old patient with cone-rod dystrophy]

"I really like the table of contents at the top as I can go straight into any topic that interests me." [P2, 52-year-old unaffected mother of a patient with type II Usher syndrome]

Readable content:

"The gene.vision page speaks to a patient or a family member. I think it addresses the issues they want to know about. The content is not too detailed, digestible and easy to understand." [P8, 45-year-old unaffected individual]

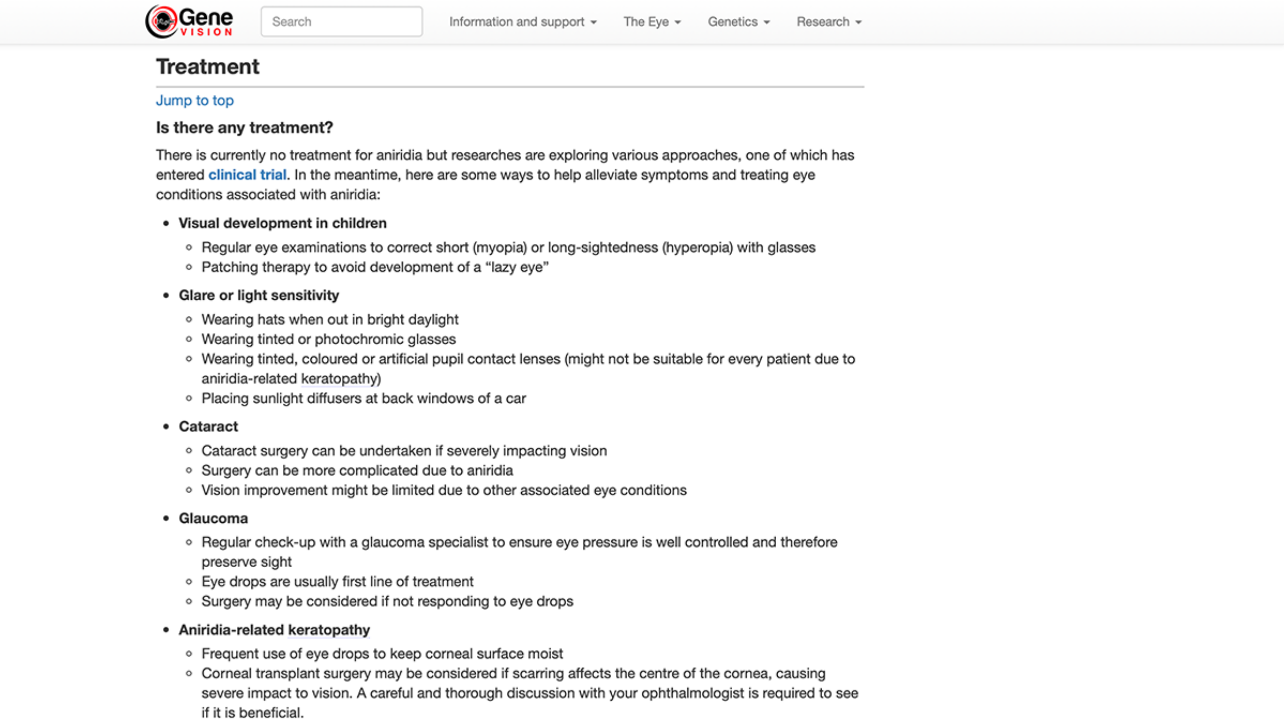
"The language used in gene.vision is good and at a good level, simple enough if you have just been diagnosed. It works well and simpler than the Wikipedia page with Voiceover." [P3, 41-year-old patient with retinitis pigmentosa]

"Gene.vision is easier to read and have a softer tone than Wikipedia. The layout is more user friendly as well." [P6, 37-year-old patient with retinitis pigmentosa]

Although unrelated to accessibility, our participants pointed out that content quality was equally important to attract users. With respect to the prototype, they felt that content related to research, support available, practical advice, and links to charities were incentives for repeated visits. They also suggested that knowing the information is coming from a reliable source will make the website more trustworthy. Quotes from participants that illustrate this point are as follows:

> *Reading stuff that is tough or difficult to absorb, it is good to have links to support. I also like the practical advice section with further links to topics that are important to me as a mother.* [P2, 52-year-old unaffected mother of a patient with type II Usher syndrome]
>
> *The content is pitched about right with links to more scientific content or papers available for doctors or patients who wanted to find out even more and look at the research themselves. It would be good to know that the website is definitely accurate if endorsed by an approved authority.* [An excerpt from the patient focus group]

Figure 4. Example of content arrangement using bullet points for ease of reading. The clinical trial link is not underlined here, which may confuse users with sight impairment.



Treatment

Jump to top

Is there any treatment?

There is currently no treatment for aniridia but researches are exploring various approaches, one of which has entered **clinical trial**. In the meantime, here are some ways to help alleviate symptoms and treating eye conditions associated with aniridia:

- **Visual development in children**
  - Regular eye examinations to correct short (myopia) or long-sightedness (hyperopia) with glasses
  - Patching therapy to avoid development of a "lazy eye"
- **Glare or light sensitivity**
  - Wearing hats when out in bright daylight
  - Wearing tinted or photochromic glasses
  - Wearing tinted, coloured or artificial pupil contact lenses (might not be suitable for every patient due to aniridia-related keratopathy)
  - Placing sunlight diffusers at back windows of a car
- **Cataract**
  - Cataract surgery can be undertaken if severely impacting vision
  - Surgery can be more complicated due to aniridia
  - Vision improvement might be limited due to other associated eye conditions
- **Glaucoma**
  - Regular check-up with a glaucoma specialist to ensure eye pressure is well controlled and therefore preserve sight
  - Eye drops are usually first line of treatment
  - Surgery may be considered if not responding to eye drops
- **Aniridia-related keratopathy**
  - Frequent use of eye drops to keep corneal surface moist
  - Corneal transplant surgery may be considered if scarring affects the centre of the cornea, causing severe impact to vision. A careful and thorough discussion with your ophthalmologist is required to see if it is beneficial.

On the other hand, our participants also identified some features on our website that impaired its usability not only for people with visual impairment but also for individuals with unimpaired vision (Textbox 4). It was highlighted to us that having good contrast is essential to accessibility. Certain aspects of our prototype, especially the navigation bar and the accompanying search box (Figure 1), had poor contrast and proved difficult for some participants to identify. One participant also pointed out that links should be underlined as default for easy identification (Figure 4). Furthermore, we found that implementing dynamic content such as an automatic timed *carousel* function (Multimedia Appendix 2) and large white spaces may frustrate and confuse screen reader and magnifier users. Having predictable navigation on a website is also crucial for accessibility. Most users are familiar with how popular search engines (Google, Bing, etc) work and they would expect similar functions when using other search boxes. This was proven when our participants found the search results page on our prototype confusing (Figure 2).

Textbox 4. Negative accessibility features of the website.

Poor chromatic and luminance contrast:

"The navigation bar is not easily found due to low contrast. I could not identify the search box on the navigation bar as it is very faint. When my pointer is at the genetics menu, it did not 'light up' very well to indicate that I am there. I find it difficult to tell if it is a link without hovering over the text with my mouse in gene.vision." [P4, 62-year-old patient with cone-rod dystrophy]

"The navigation bar is difficult to find. Grey is not visible, making it hard to see at first." [P8, 45-year-old unaffected individual]

"The 'for patients' and 'for doctors' tabs are not obvious due to poor contrast. The search box and search button could be more visible." [P5, 40-year-old patient with aniridia]

Dynamic content and large white spaces:

"I don't like the distraction of a carousel, especially one that I can't manually control. It doesn't allow me enough time to read its content." [P5, 40-year-old patient with aniridia]

"There are a lot of white space on the right side of the gene.vision page, magnifier users may get lost. The spacing is better in portrait mode on an iPad. It needs a more responsive layout." [P6, 37-year-old patient with retinitis pigmentosa]

Unpredictable search behavior:

It is difficult to find links to click on from the search results page. The 'Page title matches' and 'Page text matches' are confusing. [P8, 45-year-old unaffected individual]

"How would I use the top search box to get all the pages including the word aniridia? I have to go down to the 'containing' bit." [P5, 40-year-old patient with aniridia]

Clinician Feedback and Recommendations

All the health care professionals who tested our website felt that the information provided was comprehensive and useful. They particularly liked the *Quick links* function, which enabled them to skip to the section of interest instantly. Both consultant and trainee ophthalmologists found the information on current research very helpful, as the listed investigational therapies were explained in a simple yet detailed manner, enabling them to gain a basic understanding without having to read multiple papers. They also liked that they had quick access to scientific abstracts of seminal papers through the hyperlinks provided in each condition page. Finally, the *Information and support* menu and links to condition-specific charities were highly welcomed by all, regardless of discipline or subspecialty, as such queries are commonly posed by patients and their families in clinics. In addition, the involved health care professionals also provided some suggestions to further enhance the usability of our website. These and the improvements we have introduced to our current iteration are summarized in Textbox 5.

Textbox 5. Suggested features that we have added to the current version of the Gene.Vision website.

Suggestions

- Changing the *Overview* section of each condition to a table format for easy reference in a busy clinic
- Having a dedicated page explaining about clinical genetic testing so that clinicians can gain a basic understanding of the various sequencing techniques that are currently available
- Knowing the associated phenotypes of a specific gene can be helpful in directing further clinical management (ie, if there are any associated systemic features)
- Referral centers for specialist clinics and/or treatment should be listed where appropriate to improve practicality (eg, centers in the United Kingdom that administer a recently approved retinal gene therapy called Luxturna [24])

Actions

- The *Overview* section for each condition has been changed to a comprehensive table format
- A page on clinical genetic testing has been added to the site, providing a brief overview of the different types of sequencing methods and also covering the benefits, limitations and potential ethical concerns associated with genetic testing
- A feature called *gene cards* has been added (Multimedia Appendix 6), covering some of the more common genes listed on the Great Ormond Street Hospital for Children oculome panel [25]
- Specialist clinics for very rare conditions (eg, Bardet-Biedl syndrome) and referral centers for Luxturna treatment in the United Kingdom are listed in their associated pages

Discussion

Principal Findings

We evaluated the initial designs of a health information website focused on rare genetic eye disorders for patients, relatives, and health care professionals. Early development centered on meeting the user requirements of our patients, most of whom had visual impairment or blindness. We wanted to identify features that optimized or impaired accessibility for users with visual impairment to ensure that this vulnerable group was not excluded from information that could potentially benefit them and to improve overall user uptake.

From this testing session, we identified the following features that will enhance accessibility and usability for users with visual impairment:

- Consistent website layout and fonts.
- Structured information hierarchy with a clear description of links.
- Good chromatic and luminance contrast.
- Simple home page with consistent, predictable, and easy navigation.
- Readable content (appropriate to the intended audience).
- Adaptability to various assistive software and mobile devices.
- Avoidance of dynamic content and large *white spaces*.

In addition to good accessibility, content quality and reliability were highly rated as potential factors that may influence user traffic. They also suggested that having an authority endorsing a website will increase its trustworthiness.

On the basis of the above feedback, we updated our website by changing the features that negatively impacted accessibility for our participants. These changes were as follows:

- Improving the contrast of the navigation bar.
- Hyperlinks are bolded and underlined as default.
- Making the search box larger.
- Having a more conventional search behavior with an autocomplete function.
- Incorporating the *breadcrumb* functionality so that users can go back to the previous page that they were visiting.

We have kept features that our participants found useful as listed under *positive accessibility features of the website* in the Results section. The screen recording of the updated website is in Multimedia Appendix 6. The navigation bar and home page menu have also been updated after adding more content to the current version.

Limitations

Our study was limited by a small cohort of participants, none of whom were Braille keyboard users. Most of the participants were technologically adept (8 participants had at least basic digital skills) and belonged to the middle age bracket. Therefore, for future testing sessions, users with lower digital skills and a broader age spectrum inclusive of younger and older patients are recommended. This will consequently provide a wider opinion about web-based accessibility for people with visual impairment.

Comparison With Previous Work

The web-based behaviors of users with visual impairment differ significantly from those of individuals with full sight. Users with visual impairment tend to employ tactics to navigate around websites more efficiently. These include probing (getting a glimpse of a page by traversing it in a sequential fashion or by jumping between headings) [26], gambling scanning (skipping a determined number of lines until bumping into content that draws their interest) [27], and memorizing the amount of links that need to be skipped to get to the main content [28]. Some users may listen to the content sequentially from the beginning (exhaustive scanning), especially when visiting unfamiliar pages or sites [29].

Web accessibility for users with visual impairment can be improved by understanding the barriers they commonly encounter through usability testing and leveraging on the aforementioned navigational strategies. By having a consistent and well-structured information layout with clear headings, users were able to quickly orientate themselves without getting lost in a sea of texts [30]. This was further enhanced on our website by the *Quick links and Jump to top* functions where participants were able to navigate directly to the topic of interest without scrolling. The *Jump to top* function also served as a *shelter* if users lost their orientation within the page as they can go back to the beginning of the page instantly, which is a tactic some users employ when they are lost [30]. Clear description of links is also equally important for effective navigation, particularly for screen reader users [31].

Users with visual impairment may feel frustrated and give up on exploring a particular website further if they are faced with overwhelming situations such as a large number of search results or taking too many steps to complete a transaction [30]. The search behavior of our prototype is a prime example (Figure 2), where the search term *aniridia* yielded multiple results under the heading *page title matches* and *page text matches* with confusing excerpts. Hence, we have updated our search behavior to reflect conventional search engines and also incorporate an autocomplete function as many medical terms are not in common usage, which can make it challenging for lay users to spell accurately.

It is believed that many websites are not accessible to users with disability because developers fear that optimizing accessibility can be costly and may affect the attractiveness of a website to sighted users [32,33]. However, there is an abundance of evidence showing that usability problems are shared regardless of visual disability and everyone stands to benefit from good accessibility [34-36]. Participants with and without visual impairment in our testing session demanded similar features, such as having a simplistic design without any dynamic content. A simple and well-designed page has fewer elements to navigate and requires less time to load [31]. Furthermore, carousels cannot be manipulated with keyboards, which can lead to frustration among screen reader users. Other features such as easily readable fonts and good contrast are also crucial and are covered in the WCAG guidelines [7].

Another factor taken into consideration when designing the website was the increasing popularity of mainstream mobile devices such as smartphones and tablet computers, both among those with and without visual impairment [37]. Thus, having a *responsive* website skin that adapts to different screen sizes has been highlighted by our participants as an important accessibility feature. This will have a direct and significant impact on everyone, as nearly three-fourth of internet users are predicted to access the web solely through smartphones by 2025 [38].

This study also demonstrated the benefits of designing a website with a user-centric approach that was previously outlined by Abelse et al [39] and further emphasized in the design principles of the NHS Digital Service Manual [40]. By understanding the challenges and requirements of users, we can then create solutions and further refine them based on feedback. This is assisted by having various prototype versions so that multiple features can be tested in a single session. For example, we were able to quickly identify features that optimized or affected accessibility for those with visual impairment by having our participants test 2 home page prototypes.

Testing with target users can also provide developers with a more accurate idea of subjective features such as readability of the content and website navigation. Although most of the accessibility characteristics identified by our participants were covered by the WCAG guidelines, such parameters are often difficult or even impossible to assess with automated tools. Even if a website conforms to the WCAG guidelines, it may not ensure good accessibility as pointed out by Power et al [6]. Their study found that the WCAG 2.0 guideline only covered half of the user problems encountered by their cohort of screen reader users. Furthermore, they reported that user problems still existed despite the implementation of some of the recommendations. The authors suggested that enhancing website accessibility should shift from a traditional problem-based approach to gathering user data from testing sessions with real-world users, which is the approach we took.

Content quality is equally crucial in attracting users [39]. As a website aiming at patients and families affected by rare genetic eye disorders, we believed that having information on *soft* content, such as practical advice, available support, and charities were important to our target audience. This resonated well with our cohort and one of them, a mother of a patient affected by type II Usher syndrome (P2) stated that knowing where to look for support and learning about practical tips that will help her son’s daily living was more important as these were issues that “kept me up at night.” Litzkendorf et al [41] reported similar findings in their study where patients with rare diseases were interviewed about their information needs and acquisition. The main themes that were lacking in rare disease websites were current research, practical tips to cope with a condition, genetic counseling and family planning options, social and educational support, and the ability to connect with similar people.

The lack of exposure to rare diseases among clinicians has often led to delayed diagnoses and incorrect care for many patients. Recent surveys of primary and secondary care physicians in Belgium and Spain have identified a huge demand for information on rare diseases, particularly on genetic screening and counseling, investigational therapies, specialist referral centers, and locating reliable resources [42,43]. Although information portals for rare diseases such as Orphanet already exist, health care professionals who tested our website believed that Gene.Vision is highly relevant as it offers more in-depth information on genetic eye disorders and covers the available social support for individuals with visual impairment in the United Kingdom. Furthermore, as genomic sequencing becomes more accessible and affordable in the United Kingdom, nonmolecular ophthalmologists, with appropriate training and guidance, are encouraged to undertake genetic testing for their patients to aid diagnosis, management, and trial enrollment [44]. Once a molecular diagnosis is received, the *gene card* feature on our website supplements this by providing a quick overview of the associated phenotype or phenotypes and information on current research.

Conclusions

To make a website accessible to users with visual impairment, attention should be focused on making simple, well-designed pages with consistent layout and information structure, good contrast, and simple navigation, all of which will directly improve the overall user experience. Although most of these features are part of the WCAG 2.0 and 2.1 recommendations, usability testing with real-world users should be conducted as well as examination by professional testers and use of automated web accessibility evaluation tools. More generally, a website's design and content should be developed with the input of target users from the earliest stages to ensure that it meets their needs. Many of these steps can be implemented easily and will help in search engine optimization.

Acknowledgments

The authors would like to thank their participants for giving their time. They would also like to thank Molly Watt (Molly Watt Trust) for providing her expertise in accessibility to the development of their website and Kate Arkell from Retina UK, Linda Van de Sande, and Kelly Trout from the International WAGR Syndrome Association for reviewing the choroideremia and WAGR syndrome pages. There are no financial disclosures from all the authors. The project is funded by the Wellcome Trust (grant number: 205174/Z/16/Z), Retina UK, and the National Institute for Health Research Moorfields Biomedical Research Centre.

Authors' Contributions

JY and PT contributed equally to the development of the website, organization of the user testing session, and writing of this paper. JY also analyzed the data gathered from this testing session. JB advised on accessible website design, testing, and user research and reviewed the final manuscript. MM conceived the idea of the website, recruited participants, provided input to website content, and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screen recording of the Gene.Vision prototype homepage based on the Wikimedia system.
[MOV File , 8775 KB - jmir_v23i1e19151_app1.mov]

Multimedia Appendix 2
Screen recording of the Gene.Vision prototype homepage designed on WordPress.
[MP4 File (MP4 Video), 2135 KB - jmir_v23i1e19151_app2.mp4]

Multimedia Appendix 3
Screen recording demonstrating the "Quick links" and "Jump to top" functions in the condition pages of the prototype.
[MOV File , 27383 KB - jmir_v23i1e19151_app3.mov]

Multimedia Appendix 4
Survey form for participants of the usability testing session.
[PDF File (Adobe PDF File), 119 KB - jmir_v23i1e19151_app4.pdf]

Multimedia Appendix 5
Full results of the pretesting survey.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i1e19151_app5.xlsx]

Multimedia Appendix 6
Screen recording of the updated Gene.Vision website based on patient and clinician feedback from the usability testing session. The RPE65 gene card is demonstrated here.
[MOV File , 47109 KB - jmir_v23i1e19151_app6.mov]

References

1. Being disabled in Britain: a journey less equal. Equality and Human Rights Commission. 2017. URL: https://www.equalityhumanrights.com/sites/default/files/being-disabled-in-britain.pdf [accessed 2019-12-06]
2. Access and Inclusion in 2018: Consumers' experiences in communications markets. Office of Communications. 2019 Jan 14. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/132912/Access-and-Inclusion-report-2018.pdf [accessed 2019-12-06]
3. Wu Y, Lindsay S, Cable J, Jones R, Evans L, Xianghua X. Digital Media Usage of Sensory Impaired Users in Wales. Swansea University. 2018 Oct 11. URL: https://www.swansea.ac.uk/media/Digital-Media-Usage-of-Sensory-Impaired-Users-in-Wales.pdf [accessed 2019-12-06]
4. Seixas Pereira L, Archambault D. Web widgets barriers for visually impaired users. Stud Health Technol Inform 2017;242:836-842. [Medline: 28873893]
5. Royal National Institute of Blind People (RNIB). The economic impact of sight loss and blindness in the UK adult population, 2013. Deloitte Access Economic. 2014 Dec 05. URL: https://www.rnib.org.uk/sites/default/files/The%20economic%20impact%20of%20sight%20loss%20and%20blindness%20in%20the%20UK%202013.pdf [accessed 2020-03-30]
6. Power C, Freire A, Petrie H, Swallow D. Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 2012; Austin, USA p. 433-442. [doi: 10.1145/2207676.2207736]
7. Kirkpatrick A, Connor J, Campbell A, Cooper M. Web Content Accessibility Guidelines (WCAG) 2.1. World Wide Web Consortium. 2018 Jun 05. URL: https://www.w3.org/TR/WCAG21/ [accessed 2019-12-07]
8. Government Digital Service. Understanding new accessibility requirements for public sector bodies. Government of the United Kingdom. 2018 May 09. URL: https://www.gov.uk/guidance/accessibility-requirements-for-public-sector-websites-and-apps [accessed 2019-12-05]
9. Web Accessibility. European Commission. 2019. URL: https://ec.europa.eu/digital-single-market/en/web-accessibility [accessed 2019-12-05]
10. Christopherson R. 'Web Accessibility Guidelines' turn 10 but still less than 10% of sites are accessible. AbilityNet. 2018 Dec 18. URL: https://www.abilitynet.org.uk/news-blogs/web-accessibility-guidelines-turn-10-still-less-10-sites-are-accessible [accessed 2019-12-05]
11. Design and build digital services for the NHS. National Health Service (NHS). 2019. URL: https://service-manual.nhs.uk/ [accessed 2020-03-04]
12. Rare Disease UK. What is a rare disease? Genetic Alliance UK. 2018. URL: https://www.raredisease.org.uk/what-is-a-rare-disease/ [accessed 2020-03-11]
13. Pauer F, Litzkendorf S, Göbel J, Storf H, Zeidler J, Graf von der Schulenburg J. Rare diseases on the internet: an assessment of the quality of online information. J Med Internet Res 2017 Jan 18;19(1):e23 [FREE Full text] [doi: 10.2196/jmir.7056] [Medline: 28100442]
14. Lüchtenberg M, Kuhli-Hattenbach C, Sinangin Y, Ohrloff C, Schalnus R. Accessibility of health information on the internet to the visually impaired user. Ophthalmologica 2008;222(3):187-193. [doi: 10.1159/000126082] [Medline: 18497528]
15. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. BMJ Open 2014 Feb 12;4(2):e004015 [FREE Full text] [doi: 10.1136/bmjopen-2013-004015] [Medline: 24525390]
16. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020--the right to sight. Bull World Health Organ 2001;79(3):227-232 [FREE Full text] [Medline: 11285667]
17. Duran M. Accessibility in government: What we found when we tested tools on the world's least-accessible webpage. Government of the United Kingdom. 2017 Feb 24. URL: https://accessibility.blog.gov.uk/2017/02/24/what-we-found-when-we-tested-tools-on-the-worlds-least-accessible-webpage/ [accessed 2020-03-04]
18. How to make digital services accessible. National Health Service. 2019. URL: https://service-manual.nhs.uk/accessibility/how-to-make-digital-services-accessible [accessed 2020-03-04]
19. Yeong JL, Thomas P, Buller J, Moosajee M. A resource for patients and doctors about rare genetic eye disorders. Gene Vision. 2019 Nov 11. URL: https://dev.gene.vision/index.php/Main_Page
20. United Kingdom Government Digital Service, United Kingdom Government Cabinet Office. Policy paper: Government Digital Inclusion Strategy. Government of the United Kingdom. 2014. URL: https://www.gov.uk/government/publications/government-digital-inclusion-strategy/government-digital-inclusion-strategy [accessed 2020-02-01]
21. Heilman JM, West AG. Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. J Med Internet Res 2015;17(3):e62 [FREE Full text] [doi: 10.2196/jmir.4069] [Medline: 25739399]

22. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]
23. Wikipedia:Manual of Style/Accessibility. Wikipedia. 2020. URL: https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Accessibility [accessed 2020-07-28]
24. Russell S, Bennett J, Wellman JA, Chung DC, Yu Z, Tillman A, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. Lancet 2017 Aug 26;390(10097):849-860 [FREE Full text] [doi: 10.1016/S0140-6736(17)31868-8] [Medline: 28712537]
25. Patel A, Hayward JD, Tailor V, Nyanhete R, Ahlfors H, Gabriel C, et al. The Oculome panel test: next-generation sequencing to diagnose a diverse range of genetic developmental eye disorders. Ophthalmology 2019 Jun;126(6):888-907. [doi: 10.1016/j.ophtha.2018.12.050] [Medline: 30653986]
26. Goble C, Harper S, Stevens R. The travails of visually impaired web travellers. In: Proceedings of the eleventh ACM on Hypertext and hypermedia. 2000 Presented at: The Eleventh ACM on Hypertext and hypermedia; May 2000; San Antonio, USA p. 1-10. [doi: 10.1145/336296.336304]
27. Takagi H, Saito S, Fukuda K, Asakawa C. Analysis of navigability of web applications for improving blind usability. ACM Trans Comput Hum Interact 2007 Sep;14(3):13. [doi: 10.1145/1279700.1279703]
28. Yesilada Y, Stevens R, Harper S, Goble C. Evaluating DANTE: semantic transcoding for visually disabled users. ACM Trans Comput Hum Interact 2007 Sep;14(3):14. [doi: 10.1145/1279700.1279704]
29. WebAIM. Screen reader user survey #4. WebAIM. 2012. URL: https://webaim.org/projects/screenreadersurvey4/ [accessed 2020-07-29]
30. Vigo M, Harper S. Coping tactics employed by visually disabled users on the web. Int J Hum Comp Stud 2013 Nov;71(11):1013-1025. [doi: 10.1016/j.ijhcs.2013.08.002]
31. Kurt S. Moving toward a universally accessible web: web accessibility and education. Assist Technol 2018 Jan 16;31(4):199-208. [doi: 10.1080/10400435.2017.1414086]
32. Ellcessor E. Web Accessibility Myths as Negotiated Industrial Lore. Crit Stud Media Commun 2014 May 28;31(5):448-463. [doi: 10.1080/15295036.2014.919660]
33. Petrie H, Hamilton F, King N. Tension, what tension? Website accessibility and visual design. In: Proceedings of the 2004 international cross-disciplinary workshop on Web accessibility (W4A). 2004 Presented at: The International cross-disciplinary workshop on Web accessibility (W4A); 2004; New York, USA p. 13-18. [doi: 10.1145/990657.990660]
34. Petrie H, Kheir O. The relationship between accessibility and usability of websites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: SIGCHI conference on Human factors in computing systems; 2007; San Jose, USA p. 397-406. [doi: 10.1145/1240624.1240688]
35. Yesilada Y, Brajnik G, Vigo M, Harper S. Exploring perceptions of web accessibility: a survey approach. Behav Inf Technol 2013 Dec 13;34(2):119-134. [doi: 10.1080/0144929x.2013.848238]
36. Schmutz S, Sonderegger A, Sauer J. Implementing recommendations from web accessibility guidelines: a comparative study of nondisabled users and users with visual impairments. Hum Factors 2017 Sep;59(6):956-972. [doi: 10.1177/0018720817708397] [Medline: 28467134]
37. Martiniello N, Eisenbarth W, Lehane C, Johnson A, Wittich W. Exploring the use of smartphones and tablets among people with visual impairments: are mainstream devices replacing the use of traditional visual aids? Assist Technol 2019 Nov 07:1-12. [doi: 10.1080/10400435.2019.1682084] [Medline: 31697612]
38. Handley L. Nearly three quarters of the world will use just their smartphones to access the internet by 2025. CNBC. 2019. URL: https://www.cnbc.com/2019/01/24/smartphones-72percent-of-people-will-use-only-mobile-for-internet-by-2025.html [accessed 2020-02-20]
39. Abelse EG, Domas White M, Hahn K. A user - based design process for web sites. Internet Res 1998 Mar;8(1):39-48. [doi: 10.1108/10662249810368879]
40. Design principles. National Health Service. 2018. URL: https://service-manual.nhs.uk/design-system/design-principles [accessed 2020-03-05]
41. Litzkendorf S, Babac A, Rosenfeldt D, Schauer F, Hartz T, Lührs V, et al. Information needs of people with rare diseases-what information do patients and their relatives require. J Rare Dis Diagn Ther 2016;2(2):- [FREE Full text] [doi: 10.21767/2380-7245.100040]
42. Ramalle-Gómara E, Domínguez-Garrido E, Gómez-Eguílaz M, Marzo-Sola ME, Ramón-Trapero JL, Gil-de-Gómez J. Education and information needs for physicians about rare diseases in Spain. Orphanet J Rare Dis 2020 Jan 17;15(1):18 [FREE Full text] [doi: 10.1186/s13023-019-1285-0] [Medline: 31952528]
43. Vandeborne L, van Overbeeke E, Dooms M, De Beleyr B, Huys I. Information needs of physicians regarding the diagnosis of rare diseases: a questionnaire-based study in Belgium. Orphanet J Rare Dis 2019 May 04;14(1):99 [FREE Full text] [doi: 10.1186/s13023-019-1075-8] [Medline: 31054581]
44. Black GC, MacEwen C, Lotery AJ. The integration of genomics into clinical ophthalmic services in the UK. Eye (Lond) 2020 Jun;34(6):993-996. [doi: 10.1038/s41433-019-0704-8] [Medline: 31767964]

Abbreviations

JAWS: Job Access With Speech
NHS: National Health Service
WAGR: Wilms tumor, aniridia, genitourinary anomalies, and range of developmental delay
WCAG: Web Content Accessibility Guidelines

Edited by G Eysenbach; submitted 06.04.20; peer-reviewed by R Henderson, I MacDonald, K Liu; comments to author 29.06.20; revised version received 17.08.20; accepted 15.09.20; published 20.01.21.

Please cite as:
Yeong JL, Thomas P, Buller J, Moosajee M
A Newly Developed Web-Based Resource on Genetic Eye Disorders for Users With Visual Impairment (Gene.Vision): Usability Study
J Med Internet Res 2021;23(1):e19151
URL: http://www.jmir.org/2021/1/e19151/
doi:10.2196/19151
PMID:33470932

Original Paper

COVID-19 Misinformation Trends in Australia: Prospective Longitudinal National Survey

Kristen Pickles[1*], PhD; Erin Cvejic[1*], PhD; Brooke Nickel[1*], PhD; Tessa Copp[1*], PhD; Carissa Bonner[1*], PhD; Julie Leask[1*], PhD; Julie Ayre[1*], PhD; Carys Batcup[1*], MSc; Samuel Cornell[1*], MSc; Thomas Dakin[1*], MPH; Rachael H Dodd[1*], PhD; Jennifer M J Isautier[1*], MSc; Kirsten J McCaffery[1*], PhD

The University of Sydney, Sydney, Australia
[*]all authors contributed equally

Corresponding Author:
Kristen Pickles, PhD
The University of Sydney
Rm 127A Edward Ford Building
Sydney, 2006
Australia
Phone: 61 93512064
Email: kristen.pickles@sydney.edu.au

Abstract

Background: Misinformation about COVID-19 is common and has been spreading rapidly across the globe through social media platforms and other information systems. Understanding what the public knows about COVID-19 and identifying beliefs based on misinformation can help shape effective public health communications to ensure efforts to reduce viral transmission are not undermined.

Objective: This study aimed to investigate the prevalence and factors associated with COVID-19 misinformation in Australia and their changes over time.

Methods: This prospective, longitudinal national survey was completed by adults (18 years and above) across April (n=4362), May (n=1882), and June (n=1369) 2020.

Results: Stronger agreement with misinformation was associated with younger age, male gender, lower education level, and language other than English spoken at home ($P<.01$ for all). After controlling for these variables, misinformation beliefs were significantly associated ($P<.001$) with lower levels of digital health literacy, perceived threat of COVID-19, confidence in government, and trust in scientific institutions. Analyses of specific government-identified misinformation revealed 3 clusters: prevention (associated with male gender and younger age), causation (associated with lower education level and greater social disadvantage), and cure (associated with younger age). Lower institutional trust and greater rejection of official government accounts were associated with stronger agreement with COVID-19 misinformation.

Conclusions: The findings of this study highlight important gaps in communication effectiveness, which must be addressed to ensure effective COVID-19 prevention.

(J Med Internet Res 2021;23(1):e23805) doi:10.2196/23805

KEYWORDS

COVID-19; coronavirus; misinformation; infodemic; myths; conspiracy; digital health; literacy; social media; trust

Introduction

False, misleading, or inaccurate health information can pose a serious risk to public health and public action [1]. Misinformation about COVID-19 is common and has spread rapidly across the globe through social media platforms and other information systems [2-5]. In February 2020, the World Health Organization’s Director-General declared the global “overabundance” of COVID-19 information an “infodemic” [6]. The term “misinfodemic” has since been coined to capture the corresponding increase in misinformation surrounding the virus [7].

Misinformation, which is typically compelling, persuasive, and emotive, spreads on social media platforms significantly farther, faster, deeper, and more broadly than factual information [8]. This is particularly true within tight-knit communities, as has been observed with the spread of vaccine misinformation among some communities in the United States, Sweden, and the Netherlands [9-12]. Common COVID-19 beliefs circulating in mainstream media include framing the pandemic as a leaked bioweapon, a consequence of 5G wireless technology, a political hoax, and that the pandemic has been made up by governments to control people. Others detail ineffective measures that individuals can take to prevent or treat the disease, such as exposing themselves to sunlight or taking vitamin C [13].

Misinformation can undermine public health efforts by shaping beliefs and attitudes, particularly if encountered within a social network, and reinforcing pre-existing values and positions [14]. Importantly, lower perceived risk or perceived efficacy of prevention behaviors and altered perception of social norms might influence individuals' willingness to follow recommendations such as voluntary testing, isolation and, potentially, vaccination [15].

Understanding what the public knows about COVID-19 and identifying beliefs based on misinformation can help shape effective public health communication to ensure effort to reduce its impact, such as debunking [16].

This paper uses data from a longitudinal cohort study of the Australian public. Our aims were to: (1) investigate the prevalence of COVID-19 misinformation beliefs in the study sample; (2) examine whether any demographic, psychosocial, and cognitive factors are associated with these beliefs; and (3) investigate how these misinformation beliefs change over time.

Methods

Data Collection

The data used in this study are from a prospective, longitudinal, national survey in Australia that aimed to evaluate variation in the public's understanding, attitudes, and implementation of COVID-19 health advice during the first lockdown period in 2020 [17,18]. A total of 4362 participants were recruited between April 17 and 24, 2020; these participants completed the baseline survey (Round 1). This survey was administered 1 month after the first measures of prevention (physical distancing and quarantine) were introduced in Australia when an increasing number of COVID-19 cases were being reported. A subset (n=3214) of this sample was invited for a longitudinal follow-up to assess changes in attitudes, beliefs, and behavior over the course of the pandemic. Of the 3214 participants, 1882 (58.5%) were invited for the Round 2 survey, which was administered from May 8 to 15, 2020 (ie, 3 weeks after the baseline or Round 1). Round 3 survey was administered to 1369 of the 3214 participants (43%) from June 5 to 12, 2020 (ie, approximately 6 weeks after the baseline survey), when restrictions in Australia showed signs of easing, and the number of new COVID-19 cases and reported community transmission had drastically reduced. Round 3 survey was administered prior to the resurgence of COVID-19 cases in some areas of Australia.

Recruitment

Participants were recruited via advertisements on social media (ie, Facebook and Instagram) and by a market research company (Dynata). We used 2 different methods for recruitment with the aim of achieving a more diverse sample. Only those participants who were recruited via social media were invited for the longitudinal follow-up.

Dynata is a market research company with access to a database of 600,000 members in Australia who are willing to be involved in online research studies. Dynata invites members to participate in a certain research study only when they meet the study's eligibility criteria. For instance, only participants who met the following eligibility criteria were included in this study: adults (ie, age 18 years or older), currently living in Australia, and ability to read and understand English.

Participants recruited via Dynata received points for completing the survey; these points could then be redeemed against gift vouchers, donations to charities, or cash. Participants recruited via social media were given the opportunity to enter into a prize draw for the chance to win one of ten $20 gift cards upon completion of each survey round.

Ethics Approval

Ethics approval was obtained from the Human Research Ethics Committee of The University of Sydney (2020/212). Participants were informed about the purpose of the study, confidentiality, and risks and benefits of participation at the beginning of the survey. Completion and submission of the online questionnaire were considered as evidence of consent.

Measures

The survey was built and administered using Qualtrics (SAP SE), an online survey platform, and it was piloted within the health literacy lab. Survey items included in each round were modified from the national longitudinal study [17] to reflect psychological, behavioral, and knowledge factors considered most relevant at that stage of restrictions. Relevant measures for this study are detailed in Table 1. Age, gender, education, language other than English (LOTE) spoken at home, and socioeconomic status were assessed in Round 1, as detailed in our previous study [17].

Table 1. Measures evaluated in this study.

Item	Description and reference (if applicable)	April	May	June
Main outcomes				
COVID-19–related misinformation beliefs[a]				
	Data on the effectiveness of vaccines is often made up[b].	✓		
	Herd immunity would be beneficial for COVID-19 and this fact is covered up.	✓	✓	
	The threat of COVID-19 is greatly exaggerated.	✓	✓	✓
	Government restrictions are stronger than is needed.	✓	✓	✓
Specific COVID-19 misinformation[c]				
	5G networks are spreading the virus.			✓
	Hot temperatures kill the virus.			✓
	Vitamin C is an effective treatment.			✓
	Ibuprofen exacerbates COVID-19.			✓
	The flu shot provides immunity to COVID-19.			✓
	Hydroxychloroquine is an effective treatment.			✓
	UV rays kill the virus.			✓
	There is a cure/vaccine for COVID-19.			✓
	Parcels from China can spread the virus.			✓
	The virus causing COVID-19 was engineered and released from a Chinese laboratory in Wuhan.			✓
Explanatory variables				
Digital health literacy[d]				
	I know what health resources are available on the internet.	✓		
	I know where to find helpful health resources on the internet.	✓		
	I know how to find helpful health resources on the internet.	✓		
	I know how to use the internet to answer my questions about health.	✓		
	I know how to use the health information I find on the internet to help me.	✓		
	I have the skills I need to evaluate the health resources I find on the internet.	✓		
	I can tell high-quality health resources from low-quality health resources on the internet.	✓		
	I feel confident in using information from the internet to make health decisions.	✓		
Perceived threat of COVID-19[e]				
	Perceived public threat of COVID-19 (scale: 1=no threat at all to 10=very serious public health threat)	✓		✓
	Perceived likelihood of personally getting sick from COVID-19 (scale: 1=not at all to 5=I definitely will)	✓		✓
Confidence in the government[f]				
	I am confident in the information about COVID-19 provided by the government.	✓		✓
	I am satisfied with the amount of information about COVID-19 provided by the government.	✓		✓
	I follow government advice on social distancing to help protect the wider community.	✓		✓
	I am concerned that government recommendations about COVID-19 are not safe, or not enough is being done.	✓		✓
Trust in institutions[g]				
	Scientists involved in developing and testing new ways to control COVID-19	✓		
	Researchers involved in tracking and predicting COVID-19 cases	✓		
	Medical institutions (general practitioners, hospitals) involved in managing COVID-19 cases	✓		
COVID-19 information sources: social media				

Item	Description and reference (if applicable)	April	May	June
	Social media reported as being used as a top-3 information source	✓		✓
Rejection of official accounts[h]				
	Much of the information we receive is wrong.			✓
	I often disagree with commonly held views about the world.			✓
	Official government accounts of events cannot be trusted.			✓
	Major events are not always what they seem.			✓

[a] Four items, adapted from validated vaccine conspiracy beliefs scale [19]; scale: 1=strongly disagree to 7=strongly agree.

[b] This question was from a validated scale and referred to vaccines in general, not a COVID-19 vaccine.

[c] Ten items, taken from Australian Government Myth busting website [13]; scale: 1=definitely false to 5=definitely true.

[d] Mean of 8 items from the eHealth Literacy Scale (eHeals) [20]; scale: 1=Strongly disagree to 5=Strongly agree.

[e] Two individual items, adapted from [21].

[f] Mean of 4 items, adapted from national Australian survey on vaccination [22]; scale: 1=strongly disagree to 7=strongly agree.

[g] Mean of 3 items adapted from [23]; scale: 1=strongly disagree to 7=strongly agree.

[h] Mean of 4 items, adapted from [24]; scale: 1=strongly disagree to 5=strongly agree.

Statistical Analysis

Analyses were conducted using Stata/IC (v16.1; StataCorp LLC). The threshold for statistical significance was set at *P*<.05. Descriptive statistics (means and SD for continuous variables, and frequency and relative frequency for categorical variables) were calculated for participant characteristics and study outcomes. To reduce the number of outcomes for analysis, misinformation beliefs at baseline were combined into a single measure using principal component analysis (PCA). Associations between the extracted misinformation component and possible explanatory variables were explored using truncated linear regression (with lower-bound truncation based on the minimum numerically possible value of the extracted misinformation component that would result from responding "strongly disagree" to all question items included in the PCA) controlling for sociodemographic factors previously shown to be associated with misinformation beliefs [17].

Changes in misinformation beliefs across study rounds were examined using linear mixed models with random intercepts by the participant and robust standard errors. These items were analyzed individually owing to changes in the items included in each round.

Dimension reduction using PCA was applied to the 10 specific COVID-19 myth items (included in Round 3 of the study). Multivariable truncated regression models (with lower-bound truncation as described above) were used to examine associations with the extracted components, using the same explanatory variables as for the analysis of misinformation beliefs from Round 1. Where survey items were repeated in Round 3 (ie, perceived threat of COVID-19, confidence in government, and use of social media as a "top-3" information source), this version of the variable was included; otherwise, the response at baseline was carried forward (ie, digital health literacy, institutional trust, and sociodemographic variables). An additional explanatory variable added in Round 3 (ie, rejection of official accounts) was also included in these models.

Results

Sample Characteristics (Cross-Sectional and Longitudinal)

Sample characteristics by each month are summarized in Table 2. When compared to national data, our sample was slightly older, included more females, had higher educational attainment, and was less likely to speak a LOTE at home.

Table 2. Sample characteristics by study round (1-3).

Sample description	Values			
	Cross-sectional	Longitudinal		
	April: Round 1 (n=4362)	April: Round 1 (n=2006)[a]	May: Round 2 (n=1882)	June: Round 3 (n=1369)
Age (years), mean (SD)	42.6 (17.4)	43.1 (16.6)	43.0 (16.6)	44.6 (16.7)
Gender, n (%)				
Male	1698 (39.9)	635 (31.7)	589 (31.3)	433 (31.6)
Female	2615 (60)	1338 (66.7)	1263 (67.1)	911 (66.5)
Not specified/other	49 (1.1)	33 (1.6)	30 (1.6)	25 (1.8)
Education, n (%)				
High school or lower	934 (21.4)	317 (15.8)	302 (16.1)	198 (14.5)
Certificate I-IV[b]	617 (14.1)	223 (11.1)	204 (10.8)	140 (10.2)
University education	2811 (64.4)	1466 (73.1)	1378 (73.1)	1031 (75.3)
Language other than English spoken at home, n (%)	274 (6.3)	75 (3.7)	70 (3.7)	51 (3.7)
Cantonese	31 (0.7)	8 (0.4)	8 (0.4)	4 (0.3)
Mandarin	28 (0.6)	12 (0.6)	11 (0.6)	2 (0.1)
Spanish	19 (0.4)	6 (0.3)	6 (0.3)	2 (0.1)
Vietnamese	15 (0.3)	6 (0.3)	5 (0.3)	4 (0.3)
Hindi	14 (0.3)	1 (<0.1)	1 (0.1)	1 (0.1)
Arabic	11 (0.3)	1 (<0.1)	1 (0.1)	0
Indonesian	10 (0.2)	4 (0.2)	4 (0.2)	1 (0.1)
Urdu	10 (0.2)	2 (0.1)	2 (0.1)	2 (0.1)
Other[c]	136 (3.1)	35 (1.7)	32 (1.7)	35 (2.6)
Socioeconomic status quintile[d], mean (SD)	3.6 (1.4)	3.7 (1.4)	3.7 (1.4)	3.7 (1.4)
Residential location, n (%)				
New South Wales	2001 (45.9)	1025 (51.1)	964 (51.2)	719 (52.5)
Victoria	788 (18.1)	323 (16.1)	303 (16.1)	201 (14.7)
Queensland	672 (15.4)	280 (14.0)	254 (13.5)	183 (13.4)
Western Australia	371 (8.5)	138 (6.9)	133 (7.1)	91 (6.6)
South Australia	238 (5.5)	93 (4.6)	89 (4.7)	64 (4.7)
Tasmania	144 (3.3)	79 (3.9)	74 (3.9)	58 (4.2)
Australian Capital Territory	120 (2.8)	62 (3.1)	59 (3.1)	49 (3.6)
Northern Territory	28 (0.6)	6 (0.3)	6 (0.3)	4 (0.3)

[a] Round 1 longitudinal sample is a subsample of those included in the cross-sectional Round 1 sample, and who responded to at least one follow-up survey. This group was recruited via social media only.

[b] Certificates I-IV are tertiary qualifications; see Australian Qualifications Framework [25].

[c] Languages other than English spoken at home with cell counts <10 at baseline.

[d] Socioeconomic Indexes for Areas and Index of Relative Socioeconomic Advantage and Disadvantage Quintiles [1-5] based on participants’ residential postcode.

Misinformation Beliefs and Associations with Sociodemographic, Cognitive, and Psychosocial Variables (Cross-Sectional Sample in April)

One month into lockdown in Australia, of the 4362 participants, 753 (17.3%) agreed that data about the effectiveness of vaccines is often made up (this survey question referred to vaccines in general, not a COVID-19 vaccine); 652 (15%) agreed that herd immunity would be beneficial for COVID-19, but this is covered-up; 603 (13.8%) agreed that the threat of COVID-19 is greatly exaggerated; and 595 (13.6%) agreed that the Australian government restrictions are stronger than required. Responses on these items were moderately correlated (pairwise *r* was between 0.36 and 0.63), with good internal consistency (Cronbach α=.78) and sufficient sampling adequacy (Kaiser-Meyer-Olkin or KMO=0.76). PCA of these items resulted in the extraction of a single component with an eigenvalue greater than 1, accounting for 60.7% of the variance (component loadings are provided in Table S1 in Multimedia Appendix 1). Estimated marginal mean values from the multivariable regression model of misinformation beliefs at baseline are provided in Table 3. Stronger agreement with misinformation beliefs was significantly associated with younger age, male gender, lower education, and primarily speaking a LOTE at home (*P*<.001 for all). After controlling for these variables, misinformation beliefs were found to be significantly associated (*P*<.001 for all) with lower levels of digital health literacy, perceived threat of COVID-19, confidence in the government, and trust in scientific institutions.

Table 3. Multivariable truncated linear regression of strength of agreement with misinformation beliefs at Round 1[a]. Higher values of the outcome indicate greater support for misinformation.

Explanatory variables	Value	Estimated marginal mean differences (95% CIs)	*P* value
Sociodemographic variables			
Age in years, mean (SD)	42.5 (17.4)	−0.023 (−0.028, −0.018)	<.001
Female gender (vs male)[b], n (%)	2568 (59.9)	−0.384 (−0.541, −0.226)	<.001
Education (vs high school or less), n (%)			<.001
Certificate I-IV[c]	609 (14.2)	0.114 (−0.133, 0.360)	.37
University education	2760 (64.4)	−0.270 (−0.459, −0.080)	.005
Language other than English spoken at home, n (%)	270 (6.3)	0.847 (0.569, 1.126)	<.001
Socioeconomic status quintile, mean (SD)	3.60 (1.40)	−0.050 (−0.105, 0.005)	.08
Additional explanatory variables			
Digital health literacy[d], mean (SD)	4.04 (0.74)	−0.250 (−0.356, −0.144)	<.001
Perceived public threat of COVID-19[e], mean (SD)	7.64 (2.17)	−0.336 (−0.372, −0.300)	<.001
Not likely to get sick, n (%)	1091 (25.5)	0.649 (0.475, 0.823)	<.001
Confidence in government[f], mean (SD)	5.15 (1.06)	−0.143 (−0.222, −0.063)	<.001
Institutional trust[g], mean (SD)	5.95 (1.06)	−0.663 (−0.738, −0.587)	<.001
Social media used as a top-3 information source, n (%)	1923 (44.9)	0.151 (−0.001, 0.307)	.06

[a] Sample for analysis comprised 4286 complete records; occasional instances of missing data for explanatory variables were not imputed due to the small proportion of missingness (76/4362, 1.8%).

[b] Marginal mean differences are not reported for gender reported as “not specified” or “other” due to small sample size, but this data was included in the regression model.

[c] Certificates I-IV are tertiary qualifications; see Australian Qualifications Framework [25].

[d] Mean of 8 items, range: 1-5.

[e] Likert scale, range: 1-10.

[f] Mean of 4 items, range: 1-7.

[g] Mean of 3 items, range: 1-7.

Changes in Misinformation Beliefs Over Time (Longitudinal Sample in April-June)

The prevalence of agreement with misinformation beliefs across the study period is shown in Figure 1, which appears to be generally consistent over time. Estimated mean values from the fixed portion of linear mixed models are presented in Table 4. A significant effect of time (*P*=.006) was identified for the misinformation belief that the threat of COVID-19 is greatly exaggerated, with pairwise contrasts showing an increase in this belief between April and May; however, this difference was not maintained in June. There was a decrease in the belief that herd immunity is beneficial for COVID-19 but is covered up between April and May (*P*<.001). No difference was observed in across the study period with regard to the strength of government restrictions belief (*P*=.41).

Figure 1. Prevalence of agreement (i.e., responding as somewhat agree (5) to strongly agree (7) on the 1 to 7 Likert scale) with misinformation beliefs by study month. Error bars indicate 95% confidence intervals.].

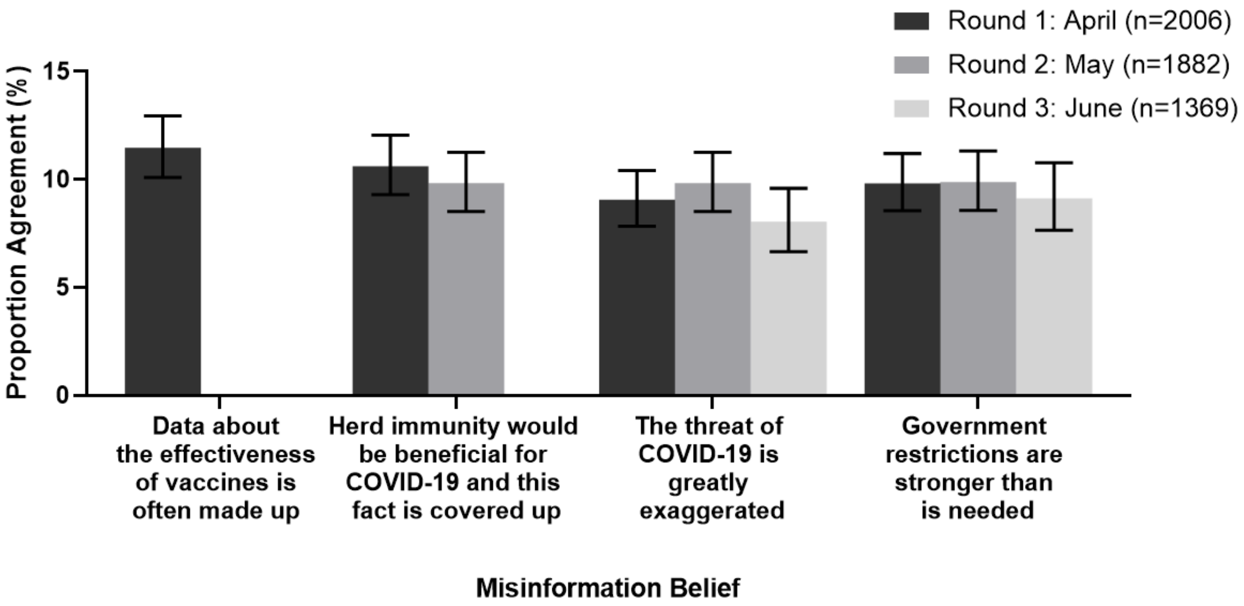


Table 4. Estimated means (95% CIs) of fixed effects from linear mixed models analyses (with random intercepts by the participant) of agreement with misinformation beliefs by study month and estimated mean differences (95% CIs) for pairwise comparisons to Round 1 (April).

Misinformation belief (scale range: 1-7)	April: Round 1 (n=2006)	May: Round 2 (n=1882)			June: Round 3 (n=1369)		
	Mean (95% CIs)	Mean (95% CIs)	Mean difference[a] (95% CIs)	*P* value	Mean (95% CIs)	Mean difference[a] (95% CIs)	*P* value
Data about the effectiveness of vaccines is often made up	2.37 (2.30, 2.44)	N/A[b]	N/A	N/A	N/A	N/A	N/A
Herd immunity would be beneficial for COVID-19 and this fact is covered up	2.52 (2.46, 2.59)	2.39 (2.32, 2.46)	−0.13 (−0.19, −0.07)	<.001	N/A	N/A	N/A
The threat of COVID-19 is greatly exaggerated	1.99 (1.93, 2.05)	2.07 (2.01, 2.14)	0.08 (0.03, 0.13)	.002	2.04 (1.98, 2.10)	0.05 (−0.01, 0.10)	.11
Government restrictions are stronger than is needed	2.14 (2.08, 2.21)	2.16 (2.09, 2.22)	0.01 (−0.04, 0.07)	.67	2.19 (2.12, 2.25)	0.04 (−0.02, 0.11)	.19

[a] Mean difference compared to Round 1 completed in April 2020.

[b] N/A: not applicable.

Specific Misinformation Beliefs and Associations With Sociodemographic, Cognitive, and Psychosocial Variables (Longitudinal Sample In June)

The level of agreement across the 10 COVID-19 misinformation items from the Australian Government website had moderate internal consistency (Cronbach α=.693) and sufficient sampling adequacy (KMO =0.761). Application of PCA (with varimax rotation) identified a 3-component solution with eigenvalues greater than 1, which cumulatively accounted for 51.15% of the variance (see Table S2 in Multimedia Appendix 1 for component loading and proportion agreement with each item). Examination of the contributing items to each component resulted in the following 3 labels:

1. Symptom management and prevention misinformation: principal component (PC)1 (explaining 18.9% of the total variance)
2. Causes and transmission misinformation: PC2 (explaining 16.7% of the total variance)
3. Immunity and cure misinformation: PC3 (explaining 15.6% of the total variance)

Regarding specific misinformation concerning symptom management and prevention, of the 1369 participants in Round 3, 301 (22%) participants agreed that hot temperatures kill the virus, 295 (21.5%) participants agreed that UV rays kill the virus, and 179 (13.1%) participants agreed that ibuprofen exacerbates COVID-19 (see Table S2 in Multimedia Appendix 1). Greater support for symptom management and prevention misinformation (PC1) was significantly associated with younger age and male gender, as well as with lower institutional trust and greater rejection of official accounts (PC1) after controlling for demographics (age, gender, education, and LOTE; see Table 5). For misinformation regarding causes and transmission, of the 1369 participants, 167 (12.2%) participants agreed that the virus causing COVID-19 was engineered and released from a Chinese laboratory in Wuhan, 57 (4.2%) participants agreed that parcels from China could spread the virus, and only 8 (0.6%) participants agreed that 5G networks are responsible for the spread of the virus. Causes and transmission misinformation (PC2) was significantly associated with less education and more social disadvantage. Greater belief in these statements was also associated with lower digital health literacy, reduced perceived public threat, reduced institutional trust, and greater rejection of official accounts after controlling for sociodemographic variables (PC2; see Table 5). Regarding misinformation about immunity and cure, of the 1369 participants in the sample, 62 (4.5%) participants agreed that vitamin C is an effective treatment, 55 (4%) participants agreed that there is a cure or vaccine for COVID-19, 32 (2.3%) participants agreed that hydroxychloroquine is an effective treatment, and 15 (1.1%) participants agreed that the flu shot provides immunity. Greater support for immunity and cure misinformation (PC3) was significantly associated with younger age. After controlling for sociodemographic factors, lower digital health literacy, reduced perceived public threat, reduced institutional trust, and greater rejection of official accounts were associated with greater belief in these statements (PC3; see Table 5).

Table 5. Multivariable truncated linear regression of the misinformation beliefs in June (Round 3)[a]. Higher values of the outcome indicate greater support for these beliefs. Data are presented as estimated marginal mean differences (95% confidence intervals) and *P* values.

Explanatory Variable	Value	Estimated marginal mean differences (95% CIs) and *P* values					
		Symptom management and prevention (PC1)	*P* value	Causes and transmission (PC2)	*P* value	Immunity and cure (PC3)	*P* value
Sociodemographic variables[b]							
Age in years, mean (SD)	44.6 (16.7)	−0.007 (−0.014, −0.001)	.03	0.005 (−0.003, 0.014)	.23	−0.021 (−0.029, −0.013)	<.001
Female gender (vs male)[c], n (%)	909 (66.5)	−0.397 (−0.610, −0.184)	<.001	0.222 (−0.087, 0.530)	.16	−0.088 (−0.341, 0.165)	.49
Education (vs high school or less), n (%)			.51		<.001		.25
Certificate I-IV[d]	140 (10.2)	0.155 (−0.245, 0.556)	.45	0.401 (−0.109, 0.912)	.13	0.266 (−0.185, 0.716)	.25
University education	1028 (75.3)	0.173 (−0.121, 0.467)	.25	−0.498 (−0.899, −0.096)	.02	−0.051 (−0.388, 0.285)	.77
Language other than English spoken at home, n (%)	51 (3.7)	−0.298 (−0.827, 0.230)	.27	0.463 (−0.254, 1.18)	.21	0.212 (−0.375, 0.799)	.48
Socioeconomic status quintile, mean (SD)	3.69 (1.39)	−0.032 (−0.104, 0.040)	.01	−0.212 (−0.313, −0.111)	<.001	−0.007 (−0.092, 0.079)	.88
Additional explanatory variables							
Digital health literacy[b,e], mean (SD)	4.18 (0.67)	0.105 (−0.046, 0.255)	.17	−0.304 (−0.512, −0.097)	.004	−0.444 (−0.618, −0.270)	<.001
Perceived public threat of COVID-19[f], mean (SD)	7.33 (2.44)	−0.027 (−0.069, 0.016)	.22	−0.074 (−0.133, −0.015)	.01	−0.057 (−0.107, −0.007)	.03
Not likely to get sick, n (%)	123 (9.0)	0.083 (−0.264, 0.429)	.64	−0.285 (−0.781, 0.211)	.26	0.133 (−0.267, 0.535)	.52
Confidence in government[g], mean (SD)	5.52 (0.94)	0.028 (−0.094, 0.149)	.67	0.117 (−0.054, 0.288)	.18	0.051 (−0.093, 0.194)	.49
Institutional trust[b,h], mean (SD)	6.15 (0.95)	−0.229 (−0.339, −0.119)	<.001	−0.599 (−0.750, −0.448)	<.001	−0.226 (−0.353, −0.099)	<.001
Social media used as a top-3 information source, n (%)	680 (49.8)	0.107 (−0.094, 0.307)	.30	0.200 (−0.087, 0.486)	.17	0.177 (−0.063, 0.417)	.15
Rejection of official accounts[i], mean (SD)	2.36 (0.83)	0.172 (0.031, 0.313)	.02	0.451 (0.245, 0.657)	<.001	0.337 (0.169, 0.506)	<.001

[a] Sample for analysis (n=1366); occasional cases of missing data for explanatory variables were not imputed due to a small proportion of missingness (3/1369, 0.2%).

[b] Values obtained in April (Round 1) and carried forward.

[c] Marginal mean differences are not reported for gender reported as "not specified" or "other" owing to small sample size but were included in the regression model.

[d] Certificates I-IV are tertiary qualifications; see the Australian Qualifications Framework [25].

[e] Mean of 8 items, range: 1-5.

[f] Likert scale, range: 1-10.

[g] Mean of 4 items, range: 1-7.

[h] Mean of 3 items, range: 1-7.

[i] Mean of 4 items, range: 1-5.

Discussion

Principal Findings

Our analysis showed lower institutional trust, lower digital health literacy, and greater rejection of official accounts were associated with a stronger agreement with COVID-19 misinformation beliefs. Misinformation was also more common among participants who primarily spoke a LOTE at home, in younger age groups, and in males. The most commonly held misinformation beliefs were concerning symptom management and prevention. We found small changes between April and May in two of the misinformation items: an increase in agreement with “COVID-19 is greatly exaggerated” and a decrease in agreement with “herd immunity is beneficial for COVID-19 but is covered up.” Despite these differences being statistically significant, they likely have little to no practical importance (ie, only a 0.08- and 0.12-unit change, respectively, on a 7-point scale). Notably, the proportion of participants agreeing with each item remained generally consistent over time during and after lockdown restrictions.

The agreement rates of COVID-19 misinformation beliefs were lower than those reported in other countries [26,27], but we note that our study was not sampled to be representative of the Australian population. An Australian poll conducted in May 2020 found relatively high support (12%-77%) for misinformation beliefs relating to the creation, spread, and prevention of the virus [28]. Interestingly, compared with the results of this poll, we found a much lower prevalence of people agreeing that 5G networks are spreading the virus. The poll found demographic patterns similar to our findings, wherein male and younger participants agreed with a range of COVID-19 misinformation beliefs more than other groups. Studies have shown that in the United States and the United Kingdom, younger people are more likely to hold conspiracy beliefs about COVID-19 [29,30]. Moreover, other studies have found that American men are more likely to agree with COVID-19 conspiracy theories than women [31].

The association between misinformation beliefs and lower education, LOTE, younger age, and male gender point toward important gaps in public health messaging to these specific groups. Our recent study highlights similar disparities in knowledge and behavior [17], as well as issues with the complexity of government health information about COVID-19. People with less education and LOTE had a poorer understanding of COVID-19 symptoms and were less frequently able to identify behaviors to prevent infection. Recently, attention has been focused on the importance of reaching people who do not speak English as their first language [32]. Our study further highlights the need for health information to be written to meet diverse health literacy requirements and targeted to specific study groups. For instance, young people and representatives of culturally and linguistically diverse groups should be involved in the design of COVID-19 messages to ensure appropriate tonality and delivery of the message. This can be achieved by testing communications with these groups, running consumer focus groups before releasing messages to the public, and ensuring representation on public health communication teams [33]. Ideally, a coproduction approach should be used to ensure targeted community messages about COVID-19 prevention are relevant and effective.

The provision of quality information online is unlikely to be a sufficient strategy to counter the influence of misinformation if digital health literacy is not accounted for. Messaging and debunking must be delivered on multiple trusted channels [34], consistent in content and style, and conveyed in local languages to ensure engagement with all communities [35]. Emerging evidence supports the idea that psychological inoculation—pre-emptively exposing people to small doses of misinformation techniques—can build resistance to false information across cultures [36]. It will be important to invest in programs teaching digital health literacy and healthy skepticism of health news, including interventions nudging people to consider the accuracy of COVID-19–related news content before sharing it further [37]. Finally, partnerships between public health authorities and trusted organizations to deliver information and correct misinformation should be utilized where possible [38]. Corrective messages are most successful when they offer a coherent explanation for how and why a belief based on misinformation is incorrect [39]. Research shows that corrective information can counter misperceptions and improve belief accuracy after an individual has been exposed to misinformation [40].

Timely, accurate, and transparent messaging is vital to gaining public trust in communication from authorities ahead of other, less credible sources [41]. Although there now is intense global interest aimed at limiting the spread of misinformation in the first place [2,36,42], this will require “a sustained and coordinated effort by independent fact-checkers, independent news media, platform companies, and public authorities to help the public understand and navigate the pandemic” [43].

Around the world and in Australia, antilockdown protests have taken place in capital cities, with protesters voicing opposition to vaccination, telecommunication towers, and COVID-19 hoax. Researchers have recently investigated the degree to which misinformation about COVID-19 is associated with people’s willingness to adhere to public health recommendations and government-enforced measures; they found that willingness decreases significantly as the strength of misbeliefs increases [44,45]; this also includes decreased intentions to avail a COVID-19 vaccination [46]. In some cases, misinformation has led to serious harm, such as the Iranian methanol poisoning episode [47]. The spread of misinformation is an ongoing area of concern as Australia and other countries worldwide continue to live with the fluctuating realities of a global pandemic. Correcting misinformation should be viewed as a vitally important science and health policy activity [48]. Importantly, the more extreme conspiracy beliefs were rare; for example, fewer than 1% of the participants in our study sample endorsed the 5G conspiracy. However, other beliefs were held by over 20% of the participants in certain demographics, indicating widespread confusion or simply outdated information spread among people, such as that regarding the use of ibuprofen.

Strengths and Limitations

The study was large and diverse but not representative of the national population. Given this, caution is needed in generalizing from these prevalence findings. The sample was recruited via an online panel and social media. The majority of participants were well educated and a low proportion were from culturally and linguistically diverse groups. Therefore, this sample may not represent the demographics of all people concerned by COVID-19 and vulnerable to misinformation, including older adults. Participants recruited via Dynata were not included in the follow-up (ie, Rounds 2 and 3) due to funding constraints. Moreover, details of the specific social media platform(s) used by the participants (eg, YouTube, Twitter, and Facebook) were not captured in our survey, but it is important to note that both good- and poor-quality information may be obtained through these channels. (Mis)information can come from various sources such as family and friends, television, radio, print media, or misinformed health care providers (including primary, allied, alternative, and complementary health sectors). The use of social media as a "top-3" information source was comparable across education categories (ie, 45% for all 3 categories); however, given the abovementioned limitation, it is unclear which platform is being used by whom.

The longitudinal design of this study enabled us to evaluate whether misinformation beliefs changed over the course of the pandemic. By design, the survey items changed across time; however, this prevented us from being able to determine longitudinal changes in the PC derived at the baseline. Finally, some of the misinformation items are likely contextual and subjective (eg, "the government restrictions are stronger than is needed"), which may have influenced the interpretation and responses of some participants.

Incorrect information about COVID-19—whether labeled as misinformation, myth, conspiracy theory, or rumor—circulates every day, and our knowledge regarding the value of various preventive interventions has progressed during the course of the pandemic. While we acknowledge that some of the misinformation items included in this survey were subject to legitimate inquiry (eg, advice recommending against the use of ibuprofen was issued by the World Health Organization early in the pandemic but then retracted), they have since been demonstrated to be scientifically incorrect, classified as misinformation, and included on myth-busting lists of leading public health institutions. The broader implication is that the groups identified in this study are more likely to agree with misinformation, including younger age, male, lower education, lower health literacy, and LOTE, may not be receiving up-to-date, evidence-based advice.

Conclusions

Misinformation can undermine public health efforts. The findings of this survey-based study highlight important gaps in communication effectiveness in the context of the COVID-19 pandemic. In efforts to prebunk and debunk misinformation, public health authorities must urgently build new partnerships with trusted, influential stakeholders and social media companies to reach the groups identified in this study. Communicators must pay close attention to ensuring that all communities can access, understand, and act on reliable COVID-19 advice.

Acknowledgments

We would like to thank all participants of this longitudinal COVID-19 survey for their ongoing participation in this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Principal component loadings from principal component analyses for COVID-19 misinformation beliefs at (1) Round 1 and (2) Round 3 with percentage agreement and descriptive statistics presented.
[DOCX File , 16 KB - jmir_v23i1e23805_app1.docx]

References

1. Tan ASL, Lee C, Chae J. Exposure to health (mis)information: lagged effects on young adults' health behaviors and potential pathways. J Commun 2015 Jul 06;65(4):674-698. [doi: 10.1111/jcom.12163]
2. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
3. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3) [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
4. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
5. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
6. Novel Coronavirus (2019-nCoV) Situation Report - 13. World Health Organization. 2020 Feb 2. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf [accessed 2020-12-16]

7. Krause NM, Freiling I, Beets B, Brossard D. Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19. Journal of Risk Research 2020 Apr 22;23(7-8):1052-1059. [doi: 10.1080/13669877.2020.1756385]
8. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
9. Feemster KA, Szipszky C. Resurgence of measles in the United States: how did we get here? Curr Opin Pediatr 2020 Feb;32(1):139-144. [doi: 10.1097/MOP.0000000000000845] [Medline: 31790030]
10. Hall V, Banerjee E, Kenyon C, Strain A, Griffith J, Como-Sabetti K, et al. Measles outbreak - Minnesota April-May 2017. MMWR Morb Mortal Wkly Rep 2017 Jul 14;66(27):713-717 [FREE Full text] [doi: 10.15585/mmwr.mm6627a1] [Medline: 28704350]
11. Ruijs WL, Hautvast JL, van der Velden K, de Vos S, Knippenberg H, Hulscher ME. Religious subgroups influencing vaccination coverage in the Dutch Bible belt: an ecological study. BMC Public Health 2011 Feb 14;11:102 [FREE Full text] [doi: 10.1186/1471-2458-11-102] [Medline: 21320348]
12. Jama A, Ali M, Lindstrand A, Butler R, Kulane A. Perspectives on the measles, mumps and rubella vaccination among Somali mothers in Stockholm. Int J Environ Res Public Health 2018 Nov 01;15(11):2428 [FREE Full text] [doi: 10.3390/ijerph15112428] [Medline: 30388799]
13. COVID-19 Mythbusting. Australian Government. 2020. URL: https://www.australia.gov.au/covid-19-mythbusting [accessed 2020-12-16]
14. Scheufele DA, Krause NM. Science audiences, misinformation, and fake news. Proc Natl Acad Sci U S A 2019 Apr 16;116(16):7662-7669 [FREE Full text] [doi: 10.1073/pnas.1805871115] [Medline: 30642953]
15. Sheeran P, Maki A, Montanaro E, Avishai-Yitshak A, Bryan A, Klein WMP, et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: A meta-analysis. Health Psychol 2016 Nov;35(11):1178-1188. [doi: 10.1037/hea0000387] [Medline: 27280365]
16. Lewandowsky S, Cook J, Ecker UKH. Under the Hood of The Debunking Handbook 2020: A consensus-based handbook of recommendations for correcting or preventing misinformation. Climate Social Science Network. 2020 Oct. URL: https://www.climatechangecommunication.org/wp-content/uploads/2020/10/DB2020paper.pdf
17. McCaffery K, Dodd RH, Cvejic E, Ayrek J, Batcup C, Isautier JM, et al. Health literacy and disparities in COVID-19-related knowledge, attitudes, beliefs and behaviours in Australia. Public Health Res Pract 2020 Dec 09;30(4):1-9 [FREE Full text] [doi: 10.17061/phrp30342012] [Medline: 33294907]
18. Dodd RH, Cvejic E, Bonner C, Pickles K, McCaffery KJ, Sydney Health Literacy Lab COVID-19 group. Willingness to vaccinate against COVID-19 in Australia. Lancet Infect Dis 2020 Jun 30;S1473-3099(20):30559-30554 [FREE Full text] [doi: 10.1016/S1473-3099(20)30559-4] [Medline: 32619436]
19. Shapiro GK, Holding A, Perez S, Amsel R, Rosberger Z. Validation of the vaccine conspiracy beliefs scale. Papillomavirus Res 2016 Dec;2:167-172 [FREE Full text] [doi: 10.1016/j.pvr.2016.09.001] [Medline: 29074176]
20. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
21. Wolf M, Serper M, Opsasnick L, O'Conor RM, Curtis L, Benavente JY, et al. Awareness, attitudes, and actions related to COVID-19 among adults with chronic conditions at the onset of the U.S. outbreak: a cross-sectional survey. Ann Intern Med 2020 Jul 21;173(2):100-109 [FREE Full text] [doi: 10.7326/M20-1239] [Medline: 32271861]
22. My C, Danchin M, Willaby HW, Pemberton S, Leask J. Parental attitudes, beliefs, behaviours and concerns towards childhood vaccinations in Australia: A national online survey. Aust Fam Physician 2017 Mar;46(3):145-151 [FREE Full text] [Medline: 28260278]
23. Frew PM, Murden R, Mehta CC, Chamberlain AT, Hinman AR, Nowak G, et al. Development of a US trust measure to assess and monitor parental confidence in the vaccine system. Vaccine 2019 Jan 07;37(2):325-332 [FREE Full text] [doi: 10.1016/j.vaccine.2018.09.043] [Medline: 30503657]
24. Uscinski JE, Enders AM, Klofstad C, Seelig M, Funchion J, Everett C, et al. Why do people believe COVID-19 conspiracy theories? HKS Misinfo Review 2020 Apr 28;1. [doi: 10.37016/mr-2020-015]
25. AQF qualifications. Australian Qualifications Framework. URL: https://www.aqf.edu.au/aqf-qualifications [accessed 2020-12-22]
26. Douglas KM, Uscinski JE, Sutton RM, Cichocka A, Nefes T, Ang CS, et al. Understanding Conspiracy Theories. Political Psychology 2019 Mar 20;40(S1):3-35. [doi: 10.1111/pops.12568]
27. Geldsetzer P. Knowledge and Perceptions of COVID-19 Among the General Public in the United States and the United Kingdom: A Cross-sectional Online Survey. Ann Intern Med 2020 Jul 21;173(2):157-160 [FREE Full text] [doi: 10.7326/M20-0912] [Medline: 32196071]
28. Belief in Conspiracy Theories. Essential Research. 2020 May 19. URL: https://essentialvision.com.au/belief-in-conspiracy-theories [accessed 2020-12-16]
29. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09:1-7. [doi: 10.1017/s003329172000224x]

30. Schaeffer K. Nearly three-in-ten Americans believe COVID-19 was made in a lab. Pew Research Centre. 2020 Apr 08. URL: https://www.pewresearch.org/fact-tank/2020/04/08/nearly-three-in-ten-americans-believe-covid-19-was-made-in-a-lab/ [accessed 2020-12-18]
31. Cassese EC, Farhart CE, Miller JM. Gender differences in COVID-19 conspiracy theory beliefs. Pol & Gen 2020 Jul 09:1-10. [doi: 10.1017/s1743923x20000409]
32. Grey A. Australia's multilingual communities are missing out on vital coronavirus information. ABC News: The Conversation. 2020 Jun 29. URL: https://www.abc.net.au/news/2020-06-29/coronavirus-multilingual-australia-missing-out-covid-19-info/12403510 [accessed 2020-12-18]
33. The best research is produced when researchers and communities work together. Nature 2018 Oct;562(7725):7. [doi: 10.1038/d41586-018-06855-7] [Medline: 30283119]
34. Fridman I, Lucas N, Henke D, Zigler CK. Association between public knowledge about COVID-19, trust in information sources, and adherence to social distancing: cross-sectional survey. JMIR Public Health Surveill 2020 Sep 15;6(3):e22060 [FREE Full text] [doi: 10.2196/22060] [Medline: 32930670]
35. Ghio D, Lawes-Wickwar S, Tang MY, Epton T, Howlett N, Jenkinson E, et al. What influences people's responses to public health messages for managing risks and preventing disease during public health crises? A rapid review of the evidence and recommendations. PsyArXiv Preprint posted online October 5, 2020. [doi: 10.31234/osf.io/nz7tr]
36. Roozenbeek J, van der Linden S, Nygren T. Prebunking interventions based on the psychological theory of “inoculation” can reduce susceptibility to misinformation across cultures. HKS Misinfo Review 2020 Feb 3. [doi: 10.37016//mr-2020-008]
37. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 2020 Jul;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]
38. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
39. Walter N, Murphy ST. How to unring the bell: A meta-analytic approach to correction of misinformation. Communication Monographs 2018 May 15;85(3):423-441. [doi: 10.1080/03637751.2018.1467564]
40. van der Meer TGLA, Jin Y. Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. Health Commun 2020 May;35(5):560-575. [doi: 10.1080/10410236.2019.1573295] [Medline: 30761917]
41. Legido-Quigley H, Asgari N, Teo YY, Leung GM, Oshitani H, Fukuda K, et al. Are high-performing health systems resilient against the COVID-19 epidemic? The Lancet 2020 Mar;395(10227):848-850. [doi: 10.1016/s0140-6736(20)30551-1]
42. Stephenson J. United Nations seeks to counter COVID-19 misinformation with digital first responders. JAMA Health Forum 2020 Jun 02;1(6):e200700. [doi: 10.1001/jamahealthforum.2020.0700]
43. Brennen JS, Simon F, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute. 2020 Apr 7. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2020-12-18]
44. Imhoff R, Bruder M. Speaking (un–)truth to power: conspiracy mentality as a generalised political attitude. Eur J Pers 2020 Dec 02;28(1):25-43. [doi: 10.1002/per.1930]
45. Oliver JE, Wood T. Medical conspiracy theories and health behaviors in the United States. JAMA Intern Med 2014 May;174(5):817-818. [doi: 10.1001/jamainternmed.2014.190] [Medline: 24638266]
46. Teovanović P, Lukić P, Zupan Z, Lazić A, Ninković M, Žeželj I. Irrational beliefs differentially predict adherence to guidelines and pseudoscientific practices during the COVID - 19 pandemic. Appl Cognit Psychol 2020 Dec 07. [doi: 10.1002/acp.3770]
47. Delirrad M, Mohammadi AB. New methanol poisoning outbreaks in Iran following COVID-19 pandemic. Alcohol Alcohol 2020 Jun 25;55(4):347-348 [FREE Full text] [doi: 10.1093/alcalc/agaa036] [Medline: 32400874]
48. Caulfield T. Does debunking work? Correcting COVID-19 misinformation on social media. OSF Preprint posted online May 25, 2020. [doi: 10.31219/osf.io/5uy2f]

Abbreviations

KMO: Kaiser-Meyer-Olkin
LOTE: language other than English
PC: principal component
PCA: principal component analysis

Edited by G Fagherazzi; submitted 24.08.20; peer-reviewed by F Sartor, P Phiri; comments to author 26.09.20; revised version received 22.10.20; accepted 09.12.20; published 07.01.21.

Please cite as:
Pickles K, Cvejic E, Nickel B, Copp T, Bonner C, Leask J, Ayre J, Batcup C, Cornell S, Dakin T, Dodd RH, Isautier JMJ, McCaffery KJ
COVID-19 Misinformation Trends in Australia: Prospective Longitudinal National Survey
J Med Internet Res 2021;23(1):e23805
URL: https://www.jmir.org/2021/1/e23805
doi:10.2196/23805
PMID:33302250

Original Paper

Evaluation of the Liver Disease Information in Baidu Encyclopedia and Wikipedia: Longitudinal Study

Fei Sun[1], MD; Fuchun Yang[1], MD; Shusen Zheng[1], MD, PhD

Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Corresponding Author:
Shusen Zheng, MD, PhD
Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery
The First Affiliated Hospital
Zhejiang University School of Medicine
#79 Qingchun Road
Hangzhou, 310003
China
Phone: 86 13805749805
Email: shusenzheng@zju.edu.cn

Abstract

Background: The internet has changed the way of people acquiring health information. Previous studies have shown that Wikipedia is a reasonably reliable medical resource, and it has been ranked higher than other general websites in various search engines. Baidu Encyclopedia is one of the most popular encyclopedia websites in China. However, no studies have shown the quality of the content provided in the Baidu Encyclopedia.

Objective: This study aimed to evaluate the quality of liver disease information provided by Wikipedia (in English) and Baidu Encyclopedia (in Chinese) and to perform a comparison of the quality and timeliness of the articles published in these two encyclopedias. Moreover, a 3-year follow-up study was conducted to compare if the information in both these websites was updated regularly over this period.

Methods: We searched for information on liver diseases by using the International Statistical Classification of Diseases and Related Health Problems 10th Revision Version 2016 codes on Wikipedia (in English) and Baidu Encyclopedia (in Chinese). The quality of the articles was assessed using the DISCERN instrument, which consists of 3 sections. We recorded the latest editing date of the webpages and calculated the date interval to evaluate the update timeliness of these websites.

Results: We found 22 entries on liver diseases in Baidu Encyclopedia and 15 articles in Wikipedia between September 15, 2016, and September 30, 2016, and we found 25 entries in Baidu Encyclopedia and 16 articles in Wikipedia between September 15, 2019, and September 30, 2019. In section 1 of the DISCERN instrument, the mean (SE) scores of Baidu Encyclopedia entries were significantly lower than those of Wikipedia articles. In section 2 and section 3 of the DISCERN instrument, the DISCERN scores of Baidu Encyclopedia entries were lower than those of Wikipedia articles, but the differences were not statistically significant. The total DISCERN scores of Baidu Encyclopedia entries were significantly lower than those of Wikipedia articles. The update interval of the entries in Baidu Encyclopedia was found to be significantly longer than that of the articles in Wikipedia.

Conclusions: This study shows that the quality of articles and the reliability of the research content on liver diseases in Wikipedia are better than those of the entries in Baidu Encyclopedia. However, the quality of the treatment choices provided in both Wikipedia and Baidu Encyclopedia is not satisfactory. Wikipedia is updated more frequently than Baidu Encyclopedia, thereby ensuring that the information presented has the most recent research findings. The findings of our study suggest that in order to find accurate health information, it is important to seek the help of medical professionals instead of looking for a prescription amid the confusing information provided on the internet.

(J Med Internet Res 2021;23(1):e17680) doi:10.2196/17680

KEYWORDS

Baidu Encyclopedia; Wikipedia; internet; website; liver disease; health information; DISCERN instrument; timeliness

Introduction

Over 4.5 billion internet users were reported worldwide in June 2019, with 854 million of them being in China [1,2]. Nowadays, the internet has become the primary source of information, and it has changed the way of people acquiring medical and health information. More than 50% of the internet users in the United States of America have been reported to search for web-based health care–related information [3,4]. A study in China found that 87.8% of the patients with scoliosis searched for scoliosis-related information on the internet, thereby indicating the high proportion of internet usage as a source of health information [5]. Increasing number of patients are seeking information about their diseases on the internet, but the reliability of web-based health care–related information is still questionable [6,7].

Wikipedia is a web-based encyclopedia that provides valuable web-based health information; it contains more than 5,956,750 articles in English [8-10]. Previous studies have shown that Wikipedia is a reasonably reliable medical resource and it was ranked higher than other general websites in search engines [8,11], although there are some errors in Wikipedia articles compared to peer-reviewed sources [12]. Baidu Encyclopedia, Wikipedia’s equivalent in China, contains more than 16,244,000 entries in Chinese [13]. Wikipedia and Baidu Encyclopedia are the most popular and consulted encyclopedia websites in English and Chinese, respectively [14-16]. On December 9, 2012, Baidu Encyclopedia announced the “rainbow plan,” wherein all medical entries could only be edited and revised by certified medical experts, which would improve the quality of the health information provided in Baidu Encyclopedia [17]. However, no study has yet examined the quality of the data provided in Baidu Encyclopedia.

Liver disease is among the top 10 causes of death in middle-income and high-income countries [18]. We have been engaged in hepatic surgery for many years and we are competent in the diagnosis and treatment of liver diseases. Therefore, we selected liver diseases as the object of analysis in this study. This study aimed to evaluate the quality of liver disease information provided by Wikipedia in English and Baidu Encyclopedia in Chinese and we aimed to perform a comparison of the quality of the research in these resources and the timeliness of the recent updates in the research between these 2 resources. Since the update frequency on the internet is high by day and by hour, 3 years can be considered as an extended period for evaluating the changes in the information over a long interval during follow-up. Thus, 3-year monitoring was conducted to compare whether the information on these websites was improved over this period. Our research results will help readers judge the reliability of web-based encyclopedia entries and avoid the medical problems caused by believing unreliable materials on the internet.

Methods

Data Sources

The articles analyzed in this study are available in Wikipedia [9] in English and in Baidu Encyclopedia [13] in Chinese. Data were compiled between September 15, 2016, and September 30, 2016. For the 3-year follow-up, data were compiled between September 15, 2019, and September 30, 2019. We chose 3 years for the length of the comparison because 3 years is a widely used and acceptable follow-up time [19-21].

Retrieval of Liver Disease Articles

The selection of topics was based on the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) version 2016. ICD defines almost all of the health-related conditions, and it is the diagnostic classification standard for all clinical and research purposes. In the ICD-10 version 2016, diseases of the liver are classified under Chapter XI Diseases of the digestive system, with categories K70 to K77 [22]. The Baidu Encyclopedia (in Chinese) and Wikipedia (in English) were investigated for articles on ICD-10 version 2016 codes. All ICD-10 code titles on the diseases of the liver were used for the search. Entries without search results were excluded from the study. Search terms on Baidu Encyclopedia (in Chinese) and Wikipedia (in English) are listed in the Multimedia Appendices 1 and 2, respectively. The proportion of the available search results were recorded and analyzed.

Assessment of the Quality of the Research Articles

The quality of the articles was assessed using the DISCERN instrument, which is used for judging the quality of health information on treatment choices [23]. The use of the DISCERN instrument for information in encyclopedias is controversial. Although some studies believe that the original edition of the DISCERN instrument is not suitable for evaluating Wikipedia articles [24,25], there are still many recent studies that have used the DISCERN instrument to evaluate the quality of Wikipedia articles [15,16]. Thus, the DISCERN instrument can be considered as an appropriate tool for evaluating the quality of Wikipedia articles in the absence of a better instrument.

The DISCERN instrument consists of 16 questions that are rated on a 5-point scale (1=definitely no, 5=definitely yes) (Multimedia Appendix 3). All these 16 questions are categorized into 3 sections. Section 1 (questions 1 to 8) assesses the reliability of a paper, section 2 (questions 9 to 15) focuses on the quality of the treatment information, and section 3 (question 16) evaluates the overall quality [23]. A higher DISCERN score indicates better paper quality. The overall DISCERN score ranges from 16 to 80 and the articles were categorized as very poor (16-26), poor (27-38), fair (39-50), good (51-62), and excellent (63-80) based on the scores [15,16].

All 3 authors have been engaged in hepatobiliary and pancreatic surgery for many years and are competent in the diagnosis and treatment of liver diseases. We have mastered the professional knowledge of liver diseases and can make professional evaluations. Sun and Yang used the DISCERN instrument to evaluate the articles in Baidu Encyclopedia (in Chinese) and Wikipedia (in English), respectively. The DISCERN score was determined through discussions. If no agreement was reached during the debate, Zheng was consulted to make the final judgment. Subsequently, Sun and Yang agreed on all the ratings.

Timeliness of the Updated Articles

We recorded and analyzed the last editing date of the web page. In 2016, the update interval was calculated as the date interval between the updated date and September 20, 2016. In 2019, the update interval was calculated as the date interval between the updated date and September 20, 2019. The update interval reflects the update frequency of a webpage. We compared the update interval of each paper to evaluate the update timeliness of the website.

Statistical Analysis

For the statistical analysis, means and standard errors (mean [SE]) were calculated. *P* values less than .05 were considered significant. Differences between groups were assessed using two-tailed Student *t* test with Welch correction. Statistical analyses were performed using GraphPad Prism 6 software (GraphPad Prism Software Inc).

Results

Retrieval of Liver Disease Articles

We searched Baidu Encyclopedia and Wikipedia with the entries in the ICD-10 diseases of the liver category. A total of 8 liver disease categories were classified, with categories K70 to K77. Only a portion of the entries were retrieved as articles in the search results. We found 22 entries in Baidu Encyclopedia and 15 articles in Wikipedia in 2016 and 25 entries in Baidu Encyclopedia and 16 articles in Wikipedia in 2019 (Table 1). In general, we were able to retrieve more entries in Baidu Encyclopedia than in Wikipedia.

Table 1. The proportion of the available search results in Baidu Encyclopedia and Wikipedia in 2016 and 2019.

ICD-10[a] category	Entries included in the ICD-10 classification (n=50)	Baidu Encyclopedia entries in 2016 (n=22)	Wikipedia articles in 2016 (n=15)	Baidu Encyclopedia entries in 2019 (n=25)	Wikipedia articles in 2019 (n=16)
K70 Alcoholic liver disease	6	4	3	4	3
K71 Toxic liver disease	10	1	1	1	1
K72 Hepatic failure, not elsewhere classified	3	1	1	2	1
K73 Chronic hepatitis, not elsewhere classified	5	3	1	3	1
K74 Fibrosis and cirrhosis of liver	7	4	1	5	1
K75 Other inflammatory liver diseases	7	3	2	3	2
K76 Other diseases of liver	10	6	6	7	7
K77 Liver disorders in diseases classified elsewhere	2	0	0	0	0

[a] ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision.

Assessment of the Quality of Articles

We used the DISCERN instrument to score the articles retrieved in Baidu Encyclopedia and Wikipedia and compared the section 1, section 2, section 3, and total scores (Figure 1). The results of the DISCERN instrument were comparable in 2016 and 2019. In section 1, the mean (SE) DISCERN score of Baidu Encyclopedia entries was significantly lower than that of Wikipedia articles (18.00 [SE 1.040] vs 26.60 [SE 1.359], *P*<.001 in 2016, respectively; 19.00 [SE 0.9110] vs 27.88 [SE 0.8892], *P*<.001 in 2019, respectively). In section 2 and section 3 of the DISCERN instrument, the mean DISCERN scores of Baidu Encyclopedia entries were lower than those of Wikipedia articles, but the differences were not statistically significant. Overall, the total scores of Baidu Encyclopedia entries were significantly lower than those of Wikipedia articles (33.68 [SE 2.265] vs 44.67 [SE 3.172], *P*=.009 in 2016, respectively; 34.72 [SE 1.943] vs 46.44 [SE 2.669], *P*=.001 in 2019, respectively) (Table 2).

The total DISCERN scores were categorized. In Baidu Encyclopedia in 2016, 9 entries were found to be of fair quality, 8 were of poor quality, and 8 were of very poor quality. In 2019, 10 entries were of fair quality, 11 were of poor quality, and 4 were of very poor quality. In Wikipedia in 2016, 5 articles were of good quality, 5 were of fair quality, 3 were of poor quality, and 2 were found to be of very poor quality. In 2019, 6 articles were found to be of good quality, 4 were of fair quality, and 5 were of poor quality.

Over time, the number of entries in Baidu Encyclopedia and Wikipedia showed an increase (Table 1). However, no significant differences were found on comparing the changes in the DISCERN scores of Baidu and Wikipedia entries over 3 years.

Figure 1. The DISCERN scores of Baidu Encyclopedia entries and Wikipedia articles on liver diseases in 2016 and 2019. $^{**}P<.01$, $^{****}P<.001$. BE-2016: Baidu Encyclopedia entries in 2016; BE-2019: Baidu Encyclopedia entries in 2019; W-2016: Wikipedia articles in 2016; W-2019: Wikipedia articles in 2019; ns: no significant difference.

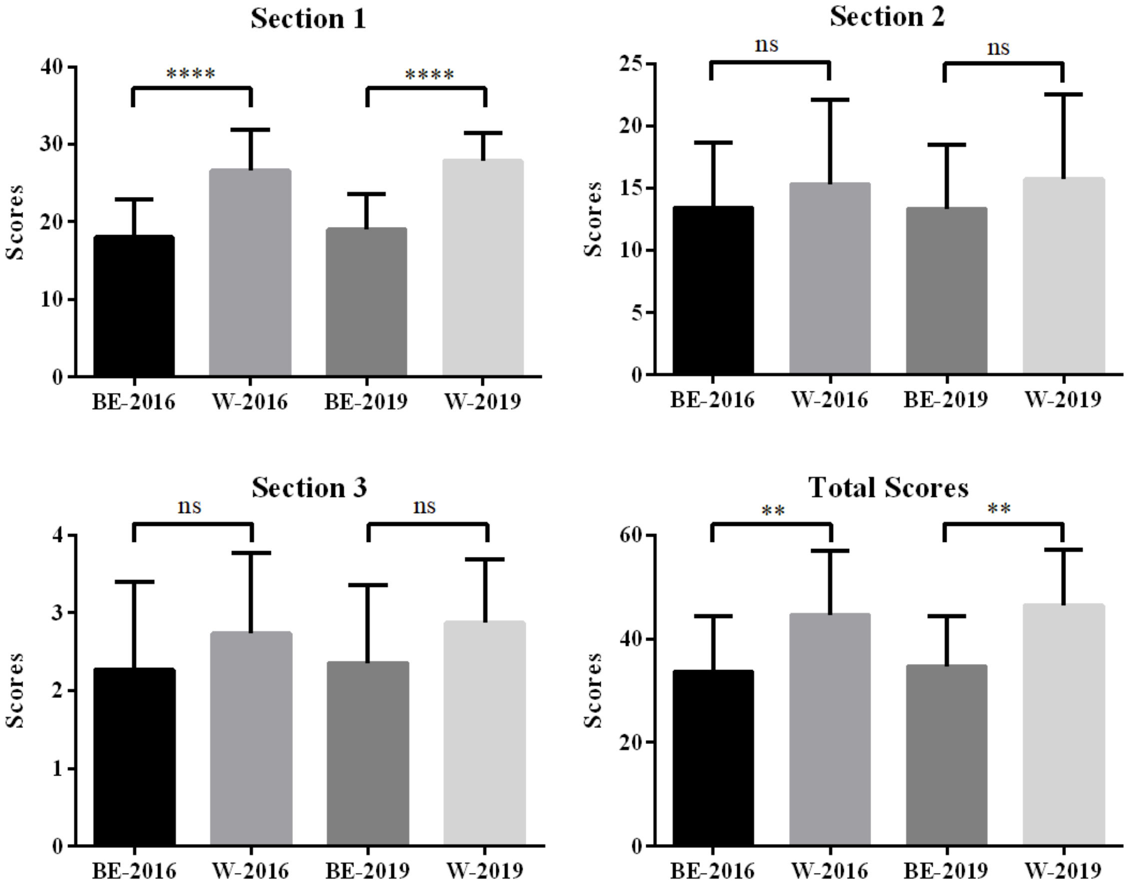


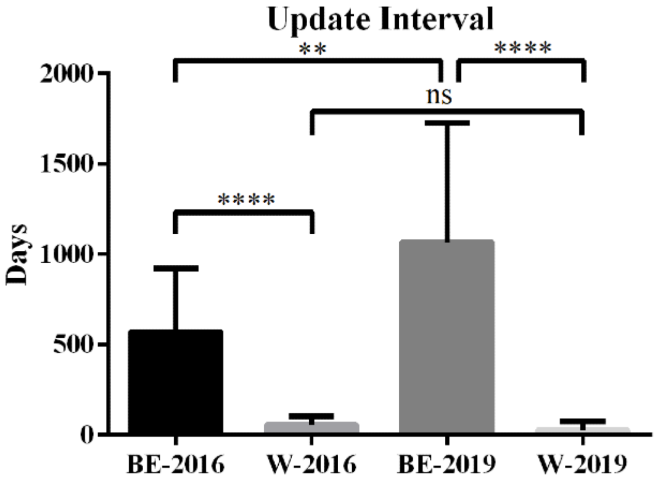
Table 2. The DISCERN scores of Baidu Encyclopedia and Wikipedia articles on liver diseases.

Year, sections of DISCERN	Baidu Encyclopedia, mean (SE)	Wikipedia, mean (SE)	*P* value
2016			
Section 1	18.00 (1.040)	26.60 (1.359)	<.001
Section 2	13.36 (1.023)	15.69 (1.717)	.25
Section 3	2.273 (0.2389)	2.733 (0.2667)	.21
Total	33.68 (2.265)	44.67 (3.172)	.009
2019			
Section 1	19.00 (0.9110)	27.88 (0.8892)	<.001
Section 2	13.36 (1.023)	15.69 (1.717)	.25
Section 3	2.360 (0.1990)	2.875 (0.2016)	.07
Total	34.72 (1.943)	46.44 (2.669)	.001

Timeliness of Articles

The update interval represents the timeliness of the updates on the website. We compared the update interval between Baidu Encyclopedia and Wikipedia (Figure 2). In 2016, the mean (SE) update interval of Baidu Encyclopedia entries was 571.6 (74.96) days and the mean (SE) update interval of Wikipedia entries was 55.93 (11.92) days. The update interval for entries in Baidu Encyclopedia was significantly longer than that for articles in Wikipedia (*P*<.001). After 3 years of follow-up, it was found that 11 of the 25 entries in Baidu Encyclopedia were recently edited before September 20, 2016, which meant that they had not been updated in the last 3 years and were in a state of loss of maintenance. In 2019, the mean (SE) update interval of Baidu Encyclopedia entries was 1067 (131.2) days and the mean (SE) update interval of Wikipedia articles was 27.44 (11.98) days. The update interval of the entries in Baidu Encyclopedia is significantly longer than that of the articles in Wikipedia (*P*<.001) (Figure 2). Therefore, the update timeliness of the articles in Wikipedia in 2016 and 2019 is significantly better than that of the entries in Baidu Encyclopedia.

Figure 2. The update interval of Baidu Encyclopedia entries and Wikipedia articles on liver diseases in 2016 and 2019. ***P*<.01, *****P*<.001. BE-2016: Baidu Encyclopedia entries in 2016; BE-2019: Baidu Encyclopedia entries in 2019; W-2016: Wikipedia articles in 2016; W-2019: Wikipedia articles in 2019; ns: no significant difference.



Discussion

Principal Findings

The main findings of this study were that the reliability of the articles on Wikipedia is better than that of the entries in Baidu Encyclopedia. However, the quality of the treatment information in both the encyclopedia websites is not satisfactory. With the increasing popularity of the internet, increasing number of patients are seeking health information and even treatment plans for their diseases on the internet [4]. However, the quality of the medical information on the internet is uneven, and a large proportion of the information is unreliable [6]. Previous research has shown that health information in a search engine is less reliable [26-28]. Wikipedia is a remarkable source of web-based health information [8]. Therefore, we analyzed the information on liver diseases in Baidu Encyclopedia in Chinese and Wikipedia in English to compare the quality of the medical information in these encyclopedia websites. The quality of the health information on the internet was studied over a 3-year follow-up period to see if it improved over time.

We evaluated the articles on liver diseases in Baidu Encyclopedia and Wikipedia in terms of quantity and quality. The Baidu Encyclopedia has more overall entries than Wikipedia; therefore, it is reasonable for Baidu Encyclopedia to have more entries in the category of liver disease than Wikipedia. In Wikipedia, terms can be easily linked to each other, making it easier for readers to build connections of knowledge [29,30]. In Baidu Encyclopedia, although there are abundant entries, the isolated entries are not linked with the other related entries, thereby making it unconducive to the formation of knowledge networks for some concepts.

We used the DISCERN instrument to evaluate the quality and reliability of the articles in Baidu Encyclopedia and Wikipedia. The DISCERN scores of Wikipedia entries were significantly higher than those of Baidu Encyclopedia entries. Consistent with that reported in previous studies, Wikipedia is a good resource of health information on the internet [8,14]. In section 1 of the DISCERN instrument, which reflects the reliability of articles, the scores for the entries in Baidu Encyclopedia were significantly lower than those for the articles in Wikipedia, indicating that the entries in Baidu Encyclopedia were less reliable than those in Wikipedia. We observed that most of the Baidu Encyclopedia entries lacked reference sources, while Wikipedia includes a detailed source for almost every entry. A high-quality citation source is an essential guarantee for the reliability of a paper [31]. However, several studies have found that Wikipedia is difficult for the general public to read and is not a reliable source for medical students [25,32]. During the 2019 follow-up, we observed the source of citations for each paper. The sources of Baidu Encyclopedia references were found to be ambiguous; they were mostly search results of Baidu academic search (an academic search engine of Baidu) and did not represent the literature cited by the entry. However, citations in Wikipedia can be explicitly identified and have original links to academic databases such as PubMed. Accurate referencing is crucial for improving the authoritativeness and credibility of a paper [31,33]. In this regard, Wikipedia was found to be better than Baidu Encyclopedia, as the basic ideas are cited and most of the references are academic papers, which guarantee scientific accuracy and reliability. In section 2 of the DISCERN instrument, which assesses the quality of the information on treatment choices, the average score of the articles in Wikipedia was found to be higher than that of the entries in Baidu Encyclopedia. However, the difference between them was not statistically significant. The scores of both Wikipedia and Baidu Encyclopedia entries were low in section 2. The quality of the information on treatment choices in both Wikipedia and Baidu Encyclopedia was not satisfactory, which is in accordance with that reported in previous studies [16,25,32]. After all, Wikipedia and Baidu Encyclopedia are not professional medical websites. Both Wikipedia and Baidu Encyclopedia have room for further improvement in providing information on treatment. Since the choice of treatment is an important target for patients to search for, adequate information should also be provided for the choice of treatment. References to international guidelines for various diseases and attracting, encouraging, and even recruiting more medical professionals to participate in the editing of the medical information provided in these websites can improve the quality of the information provided in these websites. Further, patients should be advised to seek treatment advice from more professional medical websites, and it would be more reliable to seek professional advice directly from medical personnel [34]. In section 3 of the DISCERN instrument, the score for Baidu Encyclopedia entries was lower than that for Wikipedia articles, which is consistent with the evaluation result of the total score.

In terms of the update timeliness, Wikipedia was significantly better than Baidu Encyclopedia in 2016 and 2019. During the 3-year follow-up, all Wikipedia pages were updated on time. As for Baidu Encyclopedia, among the 25 entries, 11 entries were not updated in 3 years and the longest update time reached 2564 days. Wikipedia is updated regularly and it evolves continuously while the content in Baidu Encyclopedia has been rarely maintained since its generation. The low update frequency has caused some information to be outdated, and some outdated information may even be wrong. Therefore, the timeliness of information is also an important aspect influencing the quality of medical information.

Our study uses the DISCERN instrument to evaluate the quality of Baidu Encyclopedia entries and compares the quality of these entries with that of Wikipedia articles. This is the first study to assess the quality of health information in Baidu Encyclopedia with the DISCERN instrument. Moreover, this is the first study to follow up on the quality of encyclopedia websites for 3 years. In 3 years, we found that the quality of the articles of both websites had not improved significantly.

Limitations

Our study design does have some limitations. First, the results of our research only reflect the information on liver diseases in Baidu Encyclopedia and Wikipedia on September 20, 2016, and September 20, 2019. Second, the DISCERN instrument is the only tool used to evaluate article quality in this study. Although the DISCERN instrument is widely used to evaluate website information, comprehensive use of more evaluation tools may make the results more convincible. The LIDA tool is a web-based validation instrument to measure the design and content of health information on the internet [35]. We plan to use the LIDA tool to evaluate the quality of articles in Baidu Encyclopedia and Wikipedia in our future research.

Conclusion

Although Wikipedia is not as good as the Baidu Encyclopedia in terms of the number of entries, its reliability is better than that of Baidu Encyclopedia, thereby making it more helpful to web-based health information seekers. However, the quality of the treatment choices in both Wikipedia and Baidu Encyclopedia's entries is not satisfactory. Wikipedia is updated frequently to keep the information up-to-date. The entries in Baidu Encyclopedia lack maintenance and are not updated on time; some information is outdated and some new content are lacking. In order to find accurate health information, people are advised to seek the help of medical professionals instead of looking for a prescription amid the confusing information on the internet.

Acknowledgments

Fei Sun would like to thank his wife, Wenchao Kong, for her patience, understanding, and great support during this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
International Statistical Classification of Diseases and Related Health Problems 10th Revision code search terms for liver diseases in Baidu Encyclopedia in 2019.
[PDF File (Adobe PDF File), 110 KB - jmir_v23i1e17680_app1.pdf]

Multimedia Appendix 2
International Statistical Classification of Diseases and Related Health Problems 10th Revision code search terms for liver diseases in Wikipedia in 2019.
[PDF File (Adobe PDF File), 29 KB - jmir_v23i1e17680_app2.pdf]

Multimedia Appendix 3
The DISCERN instrument.
[DOC File , 33 KB - jmir_v23i1e17680_app3.doc]

References

1. Internet World Stats. World Internet Users and 2019 Population Stats. URL: http://www.internetworldstats.com/stats.htm [accessed 2019-06-30]

2. CNNIC. China Internet Network Development State Statistic Report. URL: http://www.cac.gov.cn/2019-08/30/c_1124939590.htm [accessed 2019-08-30]
3. Jin J, Yan X, Li Y, Li Y. Int J Med Inform 2016 Feb;86:91-103. [doi: 10.1016/j.ijmedinf.2015.11.002] [Medline: 26616406]
4. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]
5. Bao H, Zhu F, Wang F, Liu Z, Bao MH, He S, et al. Scoliosis related information on the internet in China: can patients benefit from this information? PLoS One 2015;10(2):e0118289 [FREE Full text] [doi: 10.1371/journal.pone.0118289] [Medline: 25689064]
6. Jo JH, Kim JR, Kim MJ, Chung JW, Park JW. Quality and readability of online information on dental treatment for snoring and obstructive sleep apnea. Int J Med Inform 2020 Jan;133:104000. [doi: 10.1016/j.ijmedinf.2019.104000] [Medline: 31731221]
7. Baumann E, Czerwinski F, Reifegerste D. Gender-Specific Determinants and Patterns of Online Health Information Seeking: Results From a Representative German Health Survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
8. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]
9. Wikipedia. Wikipedia. URL: https://www.wikipedia.org/ [accessed 2019-09-15]
10. Farič N, Potts HWW. Motivations for contributing to health-related articles on Wikipedia: an interview study. J Med Internet Res 2014 Dec 03;16(12):e260 [FREE Full text] [doi: 10.2196/jmir.3569] [Medline: 25498308]
11. Thomas GR, Eng L, de Wolff JF, Grover SC. An evaluation of Wikipedia as a resource for patient education in nephrology. Semin Dial 2013;26(2):159-163. [doi: 10.1111/sdi.12059] [Medline: 23432369]
12. Hasty RT, Garbalosa RC, Barbato VA, Valdes PJ, Powers DW, Hernandez E, et al. Wikipedia vs peer-reviewed medical literature for information about the 10 most costly medical conditions. J Am Osteopath Assoc 2014 May;114(5):368-373. [doi: 10.7556/jaoa.2014.035] [Medline: 24778001]
13. Baidu. Baidu Encyclopedia. URL: http://baike.baidu.com/ [accessed 2019-09-15]
14. Weiner SS, Horbacewicz J, Rasberry L, Bensinger-Brody Y. Improving the Quality of Consumer Health Information on Wikipedia: Case Series. J Med Internet Res 2019 Mar 18;21(3):e12450 [FREE Full text] [doi: 10.2196/12450] [Medline: 30882357]
15. Yacob M, Lotfi S, Tang S, Jetty P. Wikipedia in Vascular Surgery Medical Education: Comparative Study. JMIR Med Educ 2020 Jun 19;6(1):e18076 [FREE Full text] [doi: 10.2196/18076] [Medline: 32417754]
16. Modiri O, Guha D, Alotaibi NM, Ibrahim GM, Lipsman N, Fallah A. Readability and quality of wikipedia pages on neurosurgical topics. Clin Neurol Neurosurg 2018 Mar;166:66-70. [doi: 10.1016/j.clineuro.2018.01.021] [Medline: 29408776]
17. Beijing Morning Post. Rainbow Plan. 2012 Dec 10. URL: http://tech.qq.com/a/20121210/000054.htm [accessed 2019-06-30]
18. Wong MCS, Huang JLW, George J, Huang J, Leung C, Eslam M, et al. The changing epidemiology of liver diseases in the Asia-Pacific region. Nat Rev Gastroenterol Hepatol 2019 Jan;16(1):57-73. [doi: 10.1038/s41575-018-0055-0] [Medline: 30158570]
19. Na X, Guo H, Zhang Y, Shen L, Wu S, Li J. Mining Open Payments Data: Analysis of Industry Payments to Thoracic Surgeons From 2014-2016. J Med Internet Res 2018 Nov 30;20(11):e11655 [FREE Full text] [doi: 10.2196/11655] [Medline: 30504119]
20. Musy SN, Endrich O, Leichtle AB, Griffiths P, Nakas CT, Simon M. Longitudinal Study of the Variation in Patient Turnover and Patient-to-Nurse Ratio: Descriptive Analysis of a Swiss University Hospital. J Med Internet Res 2020 Apr 02;22(4):e15554 [FREE Full text] [doi: 10.2196/15554] [Medline: 32238331]
21. Jabaley CS, Groff RF, Barnes TJ, Caridi-Scheible ME, Blum JM, O'Reilly-Shah VN. Sepsis information-seeking behaviors via Wikipedia between 2015 and 2018: A mixed methods retrospective observational study. PLoS One 2019;14(8):e0221596 [FREE Full text] [doi: 10.1371/journal.pone.0221596] [Medline: 31437248]
22. WHO. ICD-10 version: 2016. 2016. URL: http://apps.who.int/classifications/icd10/browse/2016/en [accessed 2016-09-30]
23. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111. [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
24. Azer SA, AlSwaidan NM, Alshwairikh LA, AlShammari JM. Accuracy and readability of cardiovascular entries on Wikipedia: are they reliable learning resources for medical students? BMJ Open 2015 Oct 06;5(10):e008187 [FREE Full text] [doi: 10.1136/bmjopen-2015-008187] [Medline: 26443650]
25. Azer SA. Is Wikipedia a reliable learning resource for medical students? Evaluating respiratory topics. Adv Physiol Educ 2015 Mar;39(1):5-14 [FREE Full text] [doi: 10.1152/advan.00110.2014] [Medline: 25727464]
26. Prasanth AS, Jayarajah U, Mohanappirian R, Seneviratne SA. Assessment of the quality of patient-oriented information over internet on testicular cancer. BMC Cancer 2018 May 02;18(1):491 [FREE Full text] [doi: 10.1186/s12885-018-4436-0] [Medline: 29716564]
27. Azer SA, Alghofaili MM, Alsultan RM, Alrumaih NS. Accuracy and Readability of Websites on Kidney and Bladder Cancers. J Cancer Educ 2018 Aug;33(4):926-944. [doi: 10.1007/s13187-017-1181-z] [Medline: 28281091]

28. Zhu X, Qiu X, Wu D, Chen S, Xiong J, Du H, et al. Reliability of information about the use of antiepileptic drugs during pregnancy from three major web search engines in China. PLoS One 2018;13(12):e0208783 [FREE Full text] [doi: 10.1371/journal.pone.0208783] [Medline: 30586373]
29. Good BM, Clarke EL, Loguercio S, Su AI. Building a biomedical semantic network in Wikipedia with Semantic Wiki Links. Database (Oxford) 2012;2012:bar060 [FREE Full text] [doi: 10.1093/database/bar060] [Medline: 22434829]
30. Lamprecht D, Lerman K, Helic D, Strohmaier M. How the structure of Wikipedia articles influences user navigation. New Rev Hypermedia Multimed 2017 Jan 02;23(1):29-50 [FREE Full text] [doi: 10.1080/13614568.2016.1179798] [Medline: 28670171]
31. Heilman JM, West AG. Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. J Med Internet Res 2015 Mar 04;17(3):e62 [FREE Full text] [doi: 10.2196/jmir.4069] [Medline: 25739399]
32. Azer SA. Evaluation of gastroenterology and hepatology articles on Wikipedia: are they suitable as learning resources for medical students? Eur J Gastroenterol Hepatol 2014 Feb;26(2):155-163. [doi: 10.1097/MEG.0000000000000003] [Medline: 24276492]
33. Luyt B, Tan D. Improving Wikipedia's credibility: References and citations in a sample of history articles. J. Am. Soc. Inf. Sci 2010;61(4):715-722. [doi: 10.1002/asi.21304]
34. Biggs TC, Jayakody N, Best K, King EV. Quality of online otolaryngology health information. J Laryngol Otol 2018 Jun;132(6):560-563. [doi: 10.1017/S0022215118000774] [Medline: 30019668]
35. Minervation. Minervation validation instrument for health care web sites, version 1.2.: Minervation; 2011. URL: http://www.minervation.com/wp-content/uploads/2011/04/Minervation-LIDA-instrument-v1-2.pdf [accessed 2019-09-30]

Abbreviations

ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision

Edited by G Eysenbach; submitted 03.01.20; peer-reviewed by L Rasberry, R Terlutter, PCI Pang; comments to author 19.03.20; revised version received 15.06.20; accepted 17.12.20; published 18.01.21.

Please cite as:
Sun F, Yang F, Zheng S
Evaluation of the Liver Disease Information in Baidu Encyclopedia and Wikipedia: Longitudinal Study
J Med Internet Res 2021;23(1):e17680
URL: http://www.jmir.org/2021/1/e17680/
doi:10.2196/17680
PMID:33459597

Original Paper

Adherence of Internet-Based Cancer Risk Assessment Tools to Best Practices in Risk Communication: Content Analysis

Erika A Waters[1], PhD, MPH; Jeremy L Foust[2], BS; Laura D Scherer[3], PhD; Amy McQueen[1], PhD; Jennifer M Taber[2], PhD

[1]Washington University School of Medicine, Saint Louis, MO, United States
[2]Kent State University, Kent, OH, United States
[3]University of Colorado School of Medicine, Denver, CO, United States

Corresponding Author:
Erika A Waters, PhD, MPH
Washington University School of Medicine
600 S Euclid Ave
Campus Box 8100
Saint Louis, MO, 63110
United States
Phone: 1 3147475705
Email: waterse@wustl.edu

Abstract

Background: Internet-based risk assessment tools offer a potential avenue for people to learn about their cancer risk and adopt risk-reducing behaviors. However, little is known about whether internet-based risk assessment tools adhere to scientific evidence for what constitutes *good* risk communication strategies. Furthermore, their quality may vary from a user experience perspective.

Objective: This study aims to understand the extent to which current best practices in risk communication have been applied to internet-based cancer risk assessment tools.

Methods: We conducted a search on August 6, 2019, to identify websites that provided personalized assessments of cancer risk or the likelihood of developing cancer. Each website (N=39) was coded according to standardized criteria and focused on 3 categories: general website characteristics, accessibility and credibility, and risk communication formats and strategies.

Results: Some best practices in risk communication were more frequently adhered to by websites. First, we found that undefined medical terminology was widespread, impeding comprehension for those with limited health literacy. For example, 90% (35/39) of websites included technical language that the general public may find difficult to understand, yet only 23% (9/39) indicated that medical professionals were their intended audience. Second, websites lacked sufficient information for users to determine the credibility of the risk assessment, making it difficult to judge the scientific validity of their risk. For instance, only 59% (23/39) of websites referenced the scientific model used to calculate the user's cancer risk. Third, practices known to foster unbiased risk comprehension, such as adding qualitative labels to quantitative numbers, were used by only 15% (6/39) of websites.

Conclusions: Limitations in risk communication strategies used by internet-based cancer risk assessment tools were common. By observing best practices, these tools could limit confusion and cultivate understanding to help people make informed decisions and motivate people to engage in risk-reducing behaviors.

(J Med Internet Res 2021;23(1):e23318)

KEYWORDS

health communication; personalized medicine; internet; risk assessment

Introduction

Background

Epidemiological risk prediction models use a patient's medical history, demographic characteristics, and/or behaviors to estimate their likelihood of experiencing a variety of clinical outcomes, such as disease incidence, progression, and survival [1]. The basic premise of many of these models is that providing people with personalized risk information will foster informed decision making (eg, engage in risk-stratified cancer screening [2,3] and take medications to prevent cancer [4]) and encourage healthy lifestyle behaviors (eg, increase physical activity engagement [5]). Furthermore, such risk models are central to a key precision medicine goal of making medical decisions based on personalized risk data [6,7]. Before the internet became widely accessible by the general public, the use of these models was restricted to clinical and research settings [4,8]. With the advancement of computing and internet technology, it has become easier for researchers to translate risk prediction models into tools that allowed for the rapid assessment of an individual's risk (hereafter, *risk assessment tool*). In 2009, a content analysis reported finding 47 websites that contained cancer-specific risk assessment tools by simply entering keywords into common search engines [9]. Since then, the number of cancer risk prediction models that could potentially be translated into risk assessment tools has dramatically increased. From 1993 to 2009, an average of 4 models were published and indexed on PubMed each year; this number increased to an average of 40 per year from January 2010 to August 2020 [10]. However, despite the increased number of risk prediction models, little is known about how internet-based risk assessment tools have (or have not) changed in the last decade.

Public-Facing Risk Assessment Tools Need to Use Good Communication Practices

The US National Cancer Institute (NCI) created the Health Information National Trends Survey (HINTS) in response to a paucity of data about the US population's cancer information needs and cancer information–seeking experiences [11]. The NCI launched the first survey in 2003, and additional iterations were administered approximately every 2 years. Data from HINTS have revealed that, from 2003 to 2019, the percentage of US residents who reported having ever sought cancer information rose from 45% to 54% [12]. More strikingly, those who used the internet as their first source of cancer information rose from approximately 46% to 75% [13,14]. Unfortunately, not only have many information seekers had negative experiences during their search, but such experiences have not improved from 2003 to 2018. Approximately 60% to 65% of HINTS respondents reported that the information was difficult to understand [15] or they were frustrated during the search process [14]. Over half of them reported concerns about the quality of the information [16].

The 2009 content analysis of internet-based cancer risk assessment tools also identified serious concerns about the quality of experiences that people might have had while searching for personalized risk assessment tools [9]. One major limitation was related to the accessibility of the information to the general public. Specifically, using medical terminology like *biopsy* without explaining what the terms meant likely limited accessibility to individuals with limited health literacy. Similarly, few tools provided an option for people who spoke a language other than English to obtain information in their native language.

The second major limitation was that only half of the websites provided information needed to evaluate their credibility, such as a citation of a peer-reviewed journal or a description of the risk prediction model used to calculate personalized estimates [17,18]. Without this information, it would be reasonable for website users to doubt the credibility of the information.

The third major category of limitations was how the actual risk information was communicated. Although most websites in the 2009 review adhered to some general principles, such as communicating numerical information as percentages or frequencies (ie, 6 in 100) and providing advice for reducing risk (eg, stop smoking), the use of other important risk communication principles was limited. Fewer than half of websites supplemented numeric estimates with qualitative labels (eg, *very high*), provided a visual display, or indicated the timeframe for which the risk was relevant (eg, 5-year risk of breast cancer) [19] (Table 1).

Table 1. Recommended risk communication formats.

Risk communication format and selected relevant citations	Why the format is important
1. Specify the duration of risk [20,21]	Time is an integral element of many risk prediction algorithms. Estimates for 1-year, 5-year, and lifetime risks can differ greatly. Consequently, specifying whether the risk estimate is applicable to the next 1 year, 5 years, 10 years, or lifetime is critical for helping people contextualize the risk and determine how much they should be concerned about it in the immediate future. For example, a 7% risk of breast cancer would be more worrisome if it was applicable to the next 5 years (in which case, it would be considered a high risk) than over one's lifetime (in that case, it would be considered a low risk). [22]
2. Provide absolute risk estimates [23,24]	Absolute risk represents the likelihood that an individual will develop a disease. It is the most basic element of risk information people need to understand their risk. Absolute risk can be communicated as numbers (eg, 12%) or as qualitative labels (eg, "very high").
3. When communicating numeric risk estimates, use percentages or simple frequencies (eg, N in 100)[a]. Do not include the 1 in N format, NNT[b], or odds ratios [25-28]	Risk comprehension is highest when risk estimates are presented as a percentage or as simple frequencies (eg, N in 100). However, both recommended formats have drawbacks. The N in 100 format can encourage people to overemphasize risk by "imagining the numerator," but the percentage format is more difficult to use while conducting complex calculations (eg, the probability of a woman having breast cancer given a positive mammogram). The 1 in N and NNT formats are very difficult to use and should be avoided.
4. When communicating numeric risk estimates, also include qualitative categories (also referred to as verbal labels; eg, high risk) [19,20,25,28,29]	Providing only numbers is problematic because of the population's low levels of numeracy (ie, the ability to use numeric information) and a lack of contextual information (eg, should a 7% lifetime risk of breast cancer be considered a high risk or a low risk?). Providing qualitative categories (also, "verbal labels") yields contextual information needed to interpret the importance of the numeric information. However, qualitative categories should be used in the absence of numeric information only with caution; people interpret qualitative categories as representing a wide range of possible probabilities (eg, "small" can mean "2%" to some people and "10%" to others). Qualitative categories can also prompt people to overestimate their risk. Providing both numeric estimates and qualitative categories yields both the detailed and contextual knowledge needed for informed decisions.
5. Frame the risk in positive and negative terms [20,21,25,30]	Framing the risk in negative terms only (eg, "Your risk of cancer is 5%") places focus only on the negative outcome and might result in exaggerated risk perceptions. Adding positive framing (eg, "This means you have a 95% chance of not getting cancer") helps participants place the risk in context.
6. If relative risk estimates are provided, then supplement it with absolute risk estimates [20,25,31]	Providing relative risk information in the absence of absolute risk information could lead people to believe that they will receive more benefit (or harm) from the action or treatment than what is possible (eg, a 50% risk reduction is much less impactful when the individual is at 4% absolute risk than at 40% absolute risk).
7. Include a visual display that depicts the individual's risk [19,20,25,28,32]	Using a visual display can increase the comprehension of risk information. However, care must be taken to avoid biasing perceptions of risk (eg, displays that focus attention on the number of people affected by a disease can exaggerate a risk as compared with displays that include information about the number of people affected and the number of people who are not affected). In addition, some visual displays are more difficult for people to understand than others (eg, pie charts).
8. Acknowledge that the risk estimate contains an element of uncertainty [25,28]	Personalized risk estimates are based on the statistical modeling of population-level data. Consequently, they always contain a level of uncertainty due to (a) limitations in the reliability or adequacy of the information used to specify the risk prediction model or (b) the randomness of future events. Informing the audience that the risk estimate is just that—an estimate—is essential to prevent them from attributing an unreasonable degree of certainty to the estimate.
9. Provide information about risk reduction strategies [23,33,34]	Understanding how to reduce risk is an essential component of understanding one's risk. Informing people how to reduce their risk is an important component of risk communication messages, particularly for individuals who have not learned risk reduction strategies previously. Providing risk information without such safety messages may undermine risk communication efforts by encouraging people to control their fear (eg, by trying to ignore the risk) rather than encouraging people to control the danger (eg, by engaging in appropriate health behaviors).
10. Consider providing comparative risk information (also called "social comparison" information) [23,28,35] but see a different perspective in studies by Fagerlin et al and Janssen et al [36,37]	Helping people understand how their risk compares to the risk of the average person of their age and sex may help them place their risk in context and evaluate its meaning and importance. For example, telling a woman that she has a 5% 5-year risk of developing breast cancer might not be meaningful unless she recognizes that it means that she is at above average risk. However, this can backfire; telling people only that they are at below-average risk might reduce motivation to engage in preventive behavior. [36]
11. Consider comparing the estimated risk to the risk of other hazards [23,28], but see different perspective in studies by Fischhoff and Covello [38,39]	Helping people understand where their risk of cancer falls in relation to other hazards such as heart disease, diabetes, or being in a car accident allows them to place the risk in context and thereby helps them determine where to invest their limited time, energy, and economic resources. However, care should be taken to avoid being perceived as condescending [38,39], lest the risk information be rejected.

[a] One controversial strategy sometimes used by risk assessment tools that aim to encourage healthy lifestyle behaviors is to communicate absolute risk information by only using qualitative (also referred to as verbal) labels [37,40] and to exclude numerical estimates [37,40]. This idea originates from

the fact that numerical estimates that public health scientists would consider high may be perceived as quite low by the public, even when a verbal label is added. For example, a person who has a 3% risk of developing lung cancer in the next 6 years would be considered high risk [41]. However, research suggests that people may consider anything below 20% risk to be *low risk*, and anything between 20% and 70% to be *medium* risk [29]. Thus, it seems likely that a person advised to quit smoking because he or she has a 3% risk of lung cancer may decide that the effort involved in quitting is not *worth* the reduction of a very small risk. However, as noted in recommendation number 4, interpretations of qualitative labels are quite variable [42] and may result in an overestimation of risk [43]. In addition, providing quantitative information may increase trust among the general public [44].

[b] NNT: number needed to treat.

Since 2009, the scientific understanding of the complexities of risk communication—including what should be considered *good* risk communication strategies—has grown. Some principles remain the same, such as avoiding the 1 in X format, acknowledging uncertainty, including visual displays, using qualitative labels to provide context for numerical risk information, and providing advice for reducing risk [19,23,25,27,45,46]. However, one piece of new knowledge (or, rather, existing knowledge made explicit) is the importance of providing risk information in a way that is likely to achieve what the communicators wish to accomplish [47,48].

One goal of providing personalized risk information is to communicate information needed to make an informed decision about preference-sensitive treatment or screening options. Examples of such decisions include mammography screening at the age of 40-49 years, prophylactic mastectomy, chemopreventive medications to reduce breast cancer risk, prostate-specific antigen screening, lung cancer screening, and colorectal cancer screening modality (eg, fecal-occult blood test, fecal immunochemical test, colonoscopy, or sigmoidoscopy). These decisions necessitate communication strategies that allow users to evaluate information in a way that, to the greatest extent possible, fosters unbiased interpretations of the risks and benefits of each option [25,48,49].

Another goal of providing personalized risk information is to motivate people to engage in risk-reducing behaviors [5,48]. Such behaviors are not preference sensitive and, barring medical contraindications, are generally recommended for entire populations. For example, many public health agencies advocate reducing cancer risk by making universal, population-wide recommendations to avoid smoking tobacco; maintain a healthy weight; be physically active; limit alcohol intake; and engage in sun protection behaviors [50]. From this perspective, providing highly detailed risk and benefit information is not as important as conveying information in a way that alerts people to their general level of risk [48] and that convinces them to engage in risk-reducing actions [47]. However, it is also critically important to provide information in such a way that is simultaneously true and persuasive. Table 1 shows the risk communication formats and strategies that can be used for both informed decision making and persuasive contexts (items 1-9) and also offers suggestions for solely persuasive purposes (items 10-11).

Study Objectives

From 2010 to 2020, many peer-reviewed articles and book chapters that describe current best practices in risk communication have been published [19,25,46,51]. In this study, we aim to understand the extent to which these practices have been applied to internet-based cancer risk assessment tools. To achieve this objective, we aim to (1) describe what risk communication formats and strategies internet-based cancer risk assessment tools use to facilitate their accessibility to the general public and demonstrate the credibility of their information, and (2) describe which risk communication formats and strategies internet-based cancer risk assessment tools use to present personalized risk information to the public.

Methods

Search Strategy and Website Selection

On August 6, 2019, we conducted an internet search to identify websites that provided people with personalized assessments of their cancer risk or the likelihood of developing cancer. One research assistant used Google and Bing (Microsoft) to search for the terms *calculate cancer risk*, *cancer risk calculator*, *estimate cancer risk*, *assess cancer risk*, and *cancer risk assessment*. Google and Bing accounted for over 80% of the world market share as of April 2020 [52]. She copied the URLs for the first 100 results for each of the 5 search terms for each of the search engines, which equaled 1000 results, into a Microsoft Excel spreadsheet. JF conducted an initial examination of these results and excluded 831 websites that were obviously not relevant (eg, risk of radiation from nuclear fallout). JF then more closely examined the remaining 169 results and excluded websites that were duplicates, redirected to another website already on the list (eg, a blog post that linked to a website already on the list), required a software download or installation of an app, had a link that no longer worked, did not provide personalized estimates of the risk of developing cancer, and/or required the user to provide medical information that must be obtained through consultation with a doctor (eg, prostate-specific antigen levels). When JF needed clarification about the appropriateness of a website, he consulted with EW; for these 13 websites, EW and JF discussed the elements of the website that were confusing and came to a decision by consensus. This process resulted in 37 websites being eligible for inclusion. Finally, we compared this list of 37 websites with the 47 websites included in the previous content analysis by Waters et al [9]. We found that 41 of the 47 websites from the 2009 analysis were no longer available. Only 6 active websites included in this review were also included in the 2009 review; 4 were identified during the search for this paper and 2 were located only by searching the 2009 paper. A total of 39 websites were included in this analysis (see Multimedia Appendix 1 for website URLs).

Measures and Coding Procedures

A total of 3 broad categories of codes were examined: general website characteristics, accessibility and credibility, and risk communication formats and strategies. The coding categories, including descriptions and/or examples of each code, were developed based on the study by Waters et al [9]. For website characteristics, we coded cancer type as assessed by the tool and the type of organization that either developed or hosted the risk assessment tool. For accessibility and credibility, we coded aspects related to accessibility to the general public, aspects of website transparency, the apparent purpose of the website (ie, whether the website appeared to be intended to provide information, recruit users, or sell products), and whether and what type of details about where to seek more information were provided. For risk communication formats and strategies, we coded whether the duration of risk was specified, whether absolute and/or comparative information was provided, whether risk was communicated qualitatively or quantitatively, the format of quantitative information (eg, percent, N in 100) if relevant, whether positive framing was used, the type of visual display used if relevant, whether or how uncertainty was acknowledged, and whether the website included information about risk reduction strategies.

Each website was independently coded by 2 of the authors (EW and JF) according to the standardized coding criteria. After independently coding the first 5 websites, the coders discussed the results and revised the coding criteria as needed. The first 5 websites were independently coded again using the updated coding scheme, and the results were discussed. Then, the remaining 34 websites were independently coded by both the coders. Finally, the codes were reviewed by both the coders to identify any discrepancies. When discrepancies occurred, the coders discussed them until a consensus was reached. In the rare case where consensus could not be reached, the coders consulted another author (JT) to make a decision. As inter-coder agreement was obtained through inter-coder checks and consensus agreement, we do not report a quantitative calculation of kappa [53].

Data Analysis

Simple frequencies were calculated and recorded for all the coding categories. We did not conduct significance testing because of the small sample sizes.

Results

General Website Characteristics

General website characteristics are summarized in Table 2. Altogether, the 39 websites provided users with up to 68 cancer risk estimates. Four websites allowed users to determine their risk of developing more than one type of cancer. In total, 16 cancer types were addressed by the calculators, with colon or colorectal, breast, and lung being the most common. The most common developer was the health care industry (eg, nonacademic-affiliated hospitals and insurance companies), which hosted just over one quarter of websites.

Table 2. General characteristics of internet-based cancer risk assessment tools (N=39 websites).

Website characteristic and description or examples of code[a]	Value, n (%)
Cancer type[b]: the type of cancer for which the tool provided personalized risk estimates	
Colon or colorectal	13 (33)
Breast	13 (33)
Lung	10 (26)
Melanoma	6 (15)
Prostate	4 (10)
Kidney	3 (8)
Ovarian	3 (8)
Pancreatic	3 (8)
Cervical	2 (5)
Stomach	2 (5)
Uterine	2 (5)
Cancer (general)	2 (5)
Other[c]	5 (13)
Website affiliation[d]: the type of organization that developed and/or hosted the tool	
Health care industry: nonacademic-affiliated hospitals or insurance companies (eg, Cigna, Merck, or Mayo Clinic)	10 (26)
Government agency: any national or international governmental agency (eg, US National Cancer Institute)	6 (15)
Educational institution: academic organization whose primary purpose is education (eg, University of California at Los Angeles)	6 (15)
Independent medical expert: risk assessment tool was developed by a specific person who purports to have relevant expertise	6 (15)
Advocacy, nonprofit: any organization that advocates and/or provides services on behalf of the public but does not fall into any other category (eg, American Cancer Society)	5 (13)
Cancer center: any organization whose primary purpose is cancer prevention and treatment (eg, Fox Chase Cancer Center)	2 (5)
General health information website: websites that do not provide services but provide health information (eg, WebMD, Yahoo! Health)	1 (3)
Other, unspecified: websites that do not fall under any existing criteria	3 (8)

[a] In most instances, the text in this column represents the explanation from the coding scheme.

[b] Categories are not mutually exclusive.

[c] The 5 *other* cancer types were blood, gastroesophageal, oral, renal, and bladder cancer.

[d] In all but one case, the organization who hosted the website also developed the website.

Accessibility and Credibility

As shown in Table 3, undefined medical terminology was widespread: close to 90% (35/39) of websites included technical language that a layperson may find difficult to understand without additional searching for information or help from a medical professional (eg, *biopsy* or *mastectomy*). However, only 23% (9/39) of websites had a statement indicating that medical professionals were the intended audience. Over half (23/39, 59%) of the websites indicated the scientific model used to calculate users' cancer risk and most provided contact information for users to ask follow-up questions about either their risk or the website (32/39, 82%). Most websites simply provided information (31/39, 80%), but a subset also appeared to attempt to recruit patients (5/39, 13% either allowed or required individuals to enter their names and contact information if they wished to be contacted by a provider) or sell products (3/39, 8% provided links to purchase materials). Most websites referred users to a doctor (34/39, 87%) and provided links to additional websites with cancer-related information (32/39, 82%).

Table 3. Accessibility and credibility of internet-based cancer risk assessment tools (N=39 websites).

Website characteristics and description or examples of code[a]	Value, n (%)
Limited accessibility for the general public: the information was not provided in a way that can be easily understood by the general public	
Undefined medical terminology: medical terminology that was not defined was included anywhere on the website (eg, ductal carcinoma in situ and biopsy)	35 (90)
Professionals only: explicit statement that the website was intended for use by medical professionals	9 (23)
Non-English option: a version of the risk assessment tool in a language other than English was provided (eg, Spanish-language version)	4 (10)
Website transparency[b]: information that allows the user to obtain information about the quality of the website	
Contact information[c]: provides information about how to contact the developers of the website	32 (82)
Indication of scientific basis of model: provides citation to peer-reviewed article about the risk prediction model; indicates which model was used (eg, Gail Model for breast cancer)	23 (59)
Last modified[c]: indicates when website was last modified	11 (28)
Underlying purpose[c]: the apparent goal of the website	
Information provision: provides risk information to website users without trying to recruit them as patients or advertise products	31 (80)
Recruit website users: provides risk information with the apparent goal of recruiting website users to use their services (eg, requires user to provide contact information before receiving risk results and asking people to fill out form with contact information if they needed a provider)	5 (13)
Sell products: provides risk information with the apparent goal of convincing users to purchase a product of theirs (eg, provide a link to purchase a book about cancer)	3 (8)
Additional resources[b,c]: details about where to seek more information	
Doctor: website advises user to seek professional assistance (eg, talk to your doctor about your risk information)	34 (87)
Links to additional websites: website provides links to locations to obtain additional cancer information (eg, to the American Cancer Society or National Cancer Institute)	32 (82)
Social media: website had a link to one or more social media pages (eg, Facebook and Twitter)	18 (46)

[a] In most instances, the text in this column represents the explanation from the coding scheme.

[b] Categories are not mutually exclusive.

[c] Code not present in the study by Waters et al [9].

Risk Communication Formats and Strategies

There was a wide variation in adherence to risk communication formats and strategies recommended for both informed decision making and encouraging behavior change. As shown in Table 4, a majority of websites followed best practice recommendations to specify a duration of risk (22/39, 56%), to provide absolute risk estimates (31/39, 80%), and to provide information about risk reduction strategies (27/39, 69%). Of the websites that provided quantitative risk information (23/39, 59%), all of them used percentages and/or natural frequencies, as recommended, and none of them used formats that are difficult for the general public to understand (eg, 1 in N, number needed to treat, and odds ratio). Very few websites followed recommendations to add qualitative labels to quantitative information (6/39, 15% of all websites and 6/23, 26% of websites that provided numbers), and only a small minority of websites framed risk information in both positive and negative terms (2/39, 5%). Consistent with recommendations, the single website that provided relative risk information also provided absolute risk information. The websites that included a visual display (20/39, 51%)—a strategy that is recommended to increase comprehension—tended to use recommended displays such as bar graphs (6/39, 15%) or icon arrays (also called stick figures; 4/39, 10%), but some websites used more difficult displays such as pie charts (3/39, 8%), line graphs (2/39, 5%), or speedometers or other less common displays (4/39, 10%). A solid majority of websites presented statements acknowledging uncertainty (23/39, 59%), as recommended. Most often, these statements indicated that the risk estimate is uncertain (19/39, 49%) or acknowledged the difficulty of estimating an individual's risk using population-based data (15/39, 39%). Few websites provided a range for the estimate or confidence intervals (2/39, 5%).

Formats and strategies that can be useful for encouraging behavior change were present in less than half of the websites: 44% (17/39) provided social comparison information and 8% (3/39) compared the risk of cancer with other hazards. Several websites (16/39, 41%) also adopted the controversial practice of providing qualitative information but not quantitative information [48] (Table 4).

Table 4. Frequency of risk communication formats and strategies used by internet-based cancer risk assessment tools (N=39 websites).

Formats and strategies	Description and/or examples of code[a]	Value, n (%)
Duration of risk	Risk information specifies the timeframe for which it applies (eg, your risk of developing melanoma in the next 5 years has been estimated as very high)	22 (56)
Absolute risk	Risk estimate includes the baseline likelihood that the individual will develop a disease; can be presented with either quantitative or numeric risk estimates or qualitative or verbal labels	31 (80)
Quantitative risk[b]	Risk estimate included numbers	23 (59)
Percent (%)	N/A[c]	23 (59)
N in 100	N/A	5 (13)
1 in N	N/A	0
Number needed to treat	N/A	0
Risk score	N/A	0
Relative risk	N/A	1 (3)
Odds	N/A	0
None provided	No quantitative risk estimate was presented	16 (41)
Qualitative risk	Risk estimate included a verbal description of risk (eg, very large)	22 (56)
Both quantitative and qualitative risk	Risk information was presented using both numbers and verbal description	6 (15)
Positive framing	Risk estimate presented as one's risk of not getting cancer (eg, this means you have a 99.2% chance of not getting cancer)	2 (5)
Visual display[b]	Whether or not there is a visual depiction of risk	20 (51)
Color[d]	Risk information is represented by at least one color	20 (51)
Bar graph[d]	Risk information presented using a bar graph	6 (15)
Icons or stick figures[d]	Risk information presented using icons (eg, smiley faces) or stick figures	4 (10)
Pie chart[d]	Risk information presented using a pie chart	3 (8)
Line graph[d]	Risk information presented using a line graph	2 (5)
Risk ladder[d]	Risk information presented in comparison with other risks on a vertical continuum	0
Table[d]	Risk information presented using a table of numbers	0
Other[d]	Risk information presented in a visual way not described above (eg, a needle on a dial pointing to where a user's risk falls)	4 (10)
Acknowledges uncertainty[b]	Website indicates there may be error or uncertainty in the risk estimate	23 (59)
Emphasizes estimate	Mentions that there is uncertainty about the accuracy of the risk estimate (eg, this is only an estimate, your actual risk may be higher or lower; the model or estimate is incomplete)	19 (49)
Individualizing population risk	Mentions the difficulty of drawing conclusions about an individual from population-based data (eg, risk is based on population, not individual; cannot determine which individuals will or will not get cancer)	15 (39)
Range or confidence intervals[d]	Provides a range of possible estimates (eg, 1.0%-2.0%)	2 (5)
Risk reduction strategies	Website provides information about how a user can reduce their risk of getting cancer (eg, quitting smoking)	27 (69)
Social comparison[b]	The website compares the user's risk with the risk of other people	17 (44)
Direct comparative[d]	The estimate is provided in a way that explicitly mentions how the user's risk compares with others (eg, your risk is higher than average)	13 (33)
Indirect comparative[d]	Separate information about the user's risk and the average person's risk is required in such a way that the user must make the comparison to determine how their risk compares with others (eg, your risk is 2%; the average woman's risk is 5%)	8 (21)
Compared with other hazards	The risk of cancer is compared with the risk of other hazards (eg, your risk of breast cancer is 10%. The risk of being involved in a car accident is 20%)	3 (8)

[a] In most instances, the text in this column represents the explanation from the coding scheme.

[b] Categories are not mutually exclusive; individual websites may have received more than one code in this category.

[c] N/A: not applicable.

[d] Code not present in the study by Waters et al [9].

Discussion

Principal Findings

Despite a decade's worth of risk communication research being conducted between our 2009 report of internet-based risk assessment tools and the investigation presented here, we found that many websites that host risk assessment tools continue to communicate information in such a way that is likely to be confusing or misunderstood. Therefore, this limits the effective translation of risk prediction models into risk assessment tools [46]. There are 3 key limitations.

First, several communication strategies limit accessibility to the general public. Although few websites state that they are only intended for medical professionals, the vast majority include medical terminology without explaining what it means. This finding is consistent with other research indicating that most internet-based risk assessment tools are difficult to use for individuals with limited health literacy [54].

Second, consistent with previous work by ourselves and others [9,18], the websites that develop and host the risk assessment tools typically do not provide sufficient information for people to determine the credibility of the tools. Although all but one website we reviewed indicated who developed it, and most websites provided a way to contact the developers in case of questions, not all websites provided information about the risk prediction model that was used to calculate their risk or provided information about when the website was last updated. Thus, users may find it difficult to determine the extent to which their personalized estimate is scientifically valid and/or personally relevant [18] and may instead rely on the credibility of the website as a heuristic to determine the credibility of the tool [55].

Third, although the use of some risk communication formats and strategies known to foster unbiased comprehension were more common than they were in 2009 (eg, using either percentages or natural frequency format to communicate quantitative risk estimates), others remain uncommon (eg, including qualitative labels when communicating quantitative risk estimates). This observation suggests that lessons learned about risk communication formats and strategies that are published in widely read journals [20,23,51] are not reaching organizations that develop and/or host websites that include risk assessment tools. This is concerning because as more risk prediction models are being developed, there is an increased likelihood of them being translated into a risk assessment tool and posted on the internet in the absence of guidance from risk communication experts [46].

One notable finding was that a considerable minority of websites used a controversial risk communication strategy that involved providing qualitative risk category labels but not quantitative risk estimates. This strategy is seen by some researchers as one that should not be used because the variability in people's interpretations of qualitative labels impedes unbiased interpretations of risk information and informed decision making [25]. However, other researchers assert that, in circumstances wherein risk information is intended to promote behavior change, only qualitative categorical information is needed, not quantitative risk estimates [48]. This latter perspective is generally consistent with health communication and promotion campaigns that inform people that engaging in a particular behavior will increase their risk, but they do not provide details about the exact extent to which their risk is increased [56]. Providing qualitative labels in the absence of quantitative data may also overcome people's tendencies to view any percentage estimate of less than 20% as a small risk [29]. This may be especially useful in situations in which researchers attempt to promote lifestyle behaviors to reduce cancer risk among healthy populations, who seldom have lifetime cancer risks of more than 20% in the absence of a high-penetrance genetic variant.

We also report 2 important differences between the work reported here and the work reported in 2009. First, 41 of the 47 websites examined in the 2009 review were no longer accessible or usable at their original URLs. Indeed, one of the websites we identified for this review was no longer accessible after 6 months. Second, this review identified several websites and tools that appeared to be developed and/or hosted by individuals who indicated they had medical or scientific expertise, but most did not appear to be affiliated with a health care institution or university. Both these findings point to the potential difficulty of verifying the validity of the risk prediction models over time. In addition, independent experts without formal affiliations to scientific institutions may have less access to collaborators with expertise in risk prediction modeling, risk communication, and behavior change than developers affiliated with governmental, educational, or health care institutions.

Limitations and Future Directions

It is possible that the initial search overlooked some relevant websites or the triage process miscategorized sites. However, our process ensured that if there were any doubts about the relevance of a website or tool, at least one other member of the coding team was consulted. We did not involve patients or members of the public in the assessment of the websites. Future research should consider doing so to increase the applicability of the results to these key end users. We also engaged in double coding for the websites included in this report; a third coder acted as a tiebreaker in particularly difficult circumstances. Unfortunately, we do not have data about how often each risk assessment tool was accessed, who accesses them, or what users do with the information they obtain from the tools. Future research should investigate this issue. We considered creating a summary score indicating how many risk communication formats and strategies were used per site, but determined that doing so was unwise because there is little research examining how many formats or strategies should be used and in which combinations. Several researchers have determined that simpler communications are either more easily understood or just as easily understood as more complex communications [37,57,58], but future research should determine which formats and strategies should be prioritized. Finally, the number of websites was too small to determine whether there were statistically significant differences between the website affiliation category and the use of risk communication formats or strategies.

Conclusions

This study found extensive variability in the extent to which internet-based risk assessment tools and the websites that host them provide information in a way that facilitates understanding among the general population. Gaps in information accessibility, in the availability of information needed to evaluate the website's credibility, and in the use of risk communication formats and strategies that foster comprehension limit the translation of risk prediction models to wider public health practice [46]. Improving the quality of communication of such tools will likely limit confusion and foster an understanding of the information among users.

Acknowledgments

This research was funded by the National Institutes of Health (R01CA190391, Principal Investigator: Waters).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Websites hosting internet-based cancer risk assessment tools.

[DOCX File , 21 KB - jmir_v23i1e23318_app1.docx]

References

1. Steyerberg E. Clinical Prediction Models. New York: Springer; 2009.
2. Walker JG, Macrae F, Winship I, Oberoi J, Saya S, Milton S, et al. The use of a risk assessment and decision support tool (CRISP) compared with usual care in general practice to increase risk-stratified colorectal cancer screening: study protocol for a randomised controlled trial. Trials 2018 Jul 25;19(1):397 [FREE Full text] [doi: 10.1186/s13063-018-2764-7] [Medline: 30045764]
3. Schapira MM, Hubbard RA, Seitz HH, Conant EF, Schnall M, Cappella JN, et al. The impact of a risk-based breast cancer screening decision aid on initiation of mammography among younger women: report of a randomized trial. MDM Policy Pract 2019;4(1):2381468318812889 [FREE Full text] [doi: 10.1177/2381468318812889] [Medline: 30729166]
4. Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Cecchini RS, Atkins JN, National Surgical Adjuvant BreastBowel Project (NSABP). Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial. JAMA 2006 Jun 21;295(23):2727-2741. [doi: 10.1001/jama.295.23.joc60074] [Medline: 16754727]
5. Fowler SL, Klein WM, Ball L, McGuire J, Colditz GA, Waters EA. Using an internet-based breast cancer risk assessment tool to improve social-cognitive precursors of physical activity. Med Decis Making 2017 Aug;37(6):657-669. [doi: 10.1177/0272989X17699835] [Medline: 28363033]
6. Bilkey GA, Burns BL, Coles EP, Mahede T, Baynam G, Nowak KJ. Optimizing precision medicine for public health. Front Public Health 2019 Mar 7;7:42 [FREE Full text] [doi: 10.3389/fpubh.2019.00042] [Medline: 30899755]
7. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795. [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
8. Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995 Feb 15;87(4):286-292. [doi: 10.1093/jnci/87.4.286]
9. Waters EA, Sullivan HW, Nelson W, Hesse BW. What is my cancer risk? How internet-based cancer risk assessment tools communicate individualized risk estimates to the public: content analysis. J Med Internet Res 2009 Jul 31;11(3):e33 [FREE Full text] [doi: 10.2196/jmir.1222] [Medline: 19674958]
10. Giardiello D, Steyerberg EW, Hauptmann M, Adank MA, Akdeniz D, Blomqvist C, et al. Prediction and clinical utility of a contralateral breast cancer risk model. Breast Cancer Res 2019 Dec 17;21(1):144 [FREE Full text] [doi: 10.1186/s13058-019-1221-1] [Medline: 31847907]
11. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004 Sep;9(5):443-60; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]
12. National Cancer Institute. Health Information National Trends Survey. 2020. URL: https://hints.cancer.gov/view-questions-topics/question-details.aspx?qid=401 [accessed 2021-01-05]

13. Huerta TR, Walker DM, Johnson T, Ford EW. A time series analysis of cancer-related information seeking: hints from the health information national trends survey (HINTS) 2003-2014. J Health Commun 2016 Sep;21(9):1031-1038. [doi: 10.1080/10810730.2016.1204381] [Medline: 27565190]
14. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online health information seeking among US adults: measuring progress toward a healthy people 2020 objective. Public Health Rep 2019 Sep 12;134(6):617-625. [doi: 10.1177/0033354919874074]
15. Health Information National Trends Survey - All HINTS Questions - Information Difficult to Find. National Cancer Institute. URL: https://hints.cancer.gov/view-questions-topics/question-details.aspx?qid=1197 [accessed 2021-01-05]
16. Health Information National Trends Survey - All HINTS Questions - Concerned about Information Quality. National Cancer Institute. 2020. URL: https://hints.cancer.gov/view-questions-topics/question-details.aspx?qid=1195 [accessed 2021-01-05]
17. Ekman A, Hall P, Litton J. Can we trust cancer information on the Internet? – A comparison of interactive cancer risk sites. Cancer Causes Control 2005 Aug;16(6):765-772. [doi: 10.1007/s10552-005-1722-3]
18. Levy AG, Sonnad SS, Kurichi JE, Sherman M, Armstrong K. Making sense of cancer risk calculators on the web. J Gen Intern Med 2008 Jan 11;23(3):229-235. [doi: 10.1007/s11606-007-0484-x]
19. Waters E, Fagerlin A, Zikmund-Fisher B. Avoiding common pitfalls of risk communication. In: Handbook of Health Decision Science. New York, NY: Springer; 2016:265-277.
20. Fagerlin A, Ubel PA, Smith DM, Zikmund-Fisher BJ. Making numbers matter: present and future research in risk communication. Am J Health Behav 2007;31 Suppl 1:S47-S56. [doi: 10.5555/ajhb.2007.31.supp.S47] [Medline: 17931136]
21. Schwartz LM, Woloshin S, Welch HG. Risk communication in clinical practice: putting cancer in context. J Natl Cancer Inst 1999 Jan 01;1999(25):124-133. [doi: 10.1093/oxfordjournals.jncimonographs.a024187]
22. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989 Dec 20;81(24):1879-1886. [doi: 10.1093/jnci/81.24.1879] [Medline: 2593165]
23. Weinstein ND. What does it mean to understand a risk? Evaluating risk comprehension. J Natl Cancer Inst Monogr 1999(25):15-20. [doi: 10.1093/oxfordjournals.jncimonographs.a024192] [Medline: 10854451]
24. Liu Y, Colditz GA, Rosner BA, Dart H, Wei E, Waters EA. Comparison of performance between a short categorized lifestyle exposure-based colon cancer risk prediction tool and a model using continuous measures. Cancer Prev Res (Phila) 2018 Dec;11(12):841-848 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-18-0196] [Medline: 30446519]
25. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PKJ, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med Inform Decis Mak 2013;13 Suppl 2:S7 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S7] [Medline: 24625237]
26. Cuite C, Weinstein ND, Emmons K, Colditz G. A test of numeric formats for communicating risk probabilities. Med Decis Making 2008;28(3):377-384. [doi: 10.1177/0272989X08315246] [Medline: 18480036]
27. Zikmund-Fisher BJ. Time to retire the 1-in-X risk format. Med Decis Making 2011;31(5):703-704. [doi: 10.1177/0272989X11418238] [Medline: 21921147]
28. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med Decis Making 2007;27(5):696-713. [doi: 10.1177/0272989X07307271] [Medline: 17873259]
29. Cameron LD, Sherman KA, Marteau TM, Brown PM. Impact of genetic risk information and type of disease on perceived risk, anticipated affect, and expected consequences of genetic tests. Health Psychol 2009 May;28(3):307-316. [doi: 10.1037/a0013947] [Medline: 19450036]
30. McNeil BJ, Pauker SG, Sox HC, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med 1982 May 27;306(21):1259-1262. [doi: 10.1056/nejm198205273062103]
31. Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. Med Decis Making 2007;27(5):638-654. [doi: 10.1177/0272989X07306783] [Medline: 17873250]
32. Garcia-Retamero R, Cokely ET. Communicating health risks with visual aids. Curr Dir Psychol Sci 2013 Sep 25;22(5):392-399. [doi: 10.1177/0963721413491570]
33. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]
34. Rogers R, Prentice-Dunn S. Protection motivation theory. In: Handbook of Health Behavior Research I: Personal and Social Determinants. New York: Plenum Press; 1997:113-132.
35. KLEIN WM. Comparative risk estimates relative to the average peer predict behavioral intentions and concern about absolute risk. Risk Dec Policy 2002;7(2):193-202. [doi: 10.1017/s1357530902000613]
36. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. If I'm better than average, then I'm ok?: Comparative information influences beliefs about risk and benefits. Patient Educ Couns 2007 Dec;69(1-3):140-144 [FREE Full text] [doi: 10.1016/j.pec.2007.08.008] [Medline: 17942271]
37. Janssen E, Ruiter RAC, Waters EA. Combining risk communication strategies to simultaneously convey the risks of four diseases associated with physical inactivity to socio-demographically diverse populations. J Behav Med 2018 Jun;41(3):318-332 [FREE Full text] [doi: 10.1007/s10865-017-9894-3] [Medline: 29027602]

38. Fischhoff B. Risk perception and communication unplugged: Twenty years of progress. Risk Anal 1995;15(2):137-145. [doi: 10.1111/j.1539-6924.1995.tb00308.x]
39. Covello VT. Risk communication, radiation, and radiological emergencies: strategies, tools, and techniques. Health Phys 2011 Nov;101(5):511-530. [doi: 10.1097/HP.0b013e3182299549] [Medline: 21979532]
40. Weinstein ND, Atwood K, Puleo E, Fletcher R, Colditz G, Emmons KM. Colon cancer: risk perceptions and risk communication. J Health Commun 2004;9(1):53-65. [doi: 10.1080/10810730490271647] [Medline: 14761833]
41. Sakoda LC, Henderson LM, Caverly TJ, Wernli KJ, Katki HA. Applying risk prediction models to optimize lung cancer screening: current knowledge, challenges, and future directions. Curr Epidemiol Rep 2017 Dec;4(4):307-320 [FREE Full text] [doi: 10.1007/s40471-017-0126-8] [Medline: 29531893]
42. Waters EA, Weinstein ND, Colditz GA, Emmons K. Formats for improving risk communication in medical tradeoff decisions. J Health Commun 2006 Mar;11(2):167-182. [doi: 10.1080/10810730500526695] [Medline: 16537286]
43. Berry DC, Knapp P, Raynor T. Expressing medicine side effects: assessing the effectiveness of absolute risk, relative risk, and number needed to harm, and the provision of baseline risk information. Patient Educ Couns 2006 Oct;63(1-2):89-96. [doi: 10.1016/j.pec.2005.09.003] [Medline: 16242904]
44. Gurmankin AD, Baron J, Armstrong K. The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. Med Decis Making 2004;24(3):265-271. [doi: 10.1177/0272989X04265482] [Medline: 15155015]
45. Politi MC, Han PKJ, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. Med Decis Making 2007;27(5):681-695. [doi: 10.1177/0272989X07307270] [Medline: 17873256]
46. Waters EA, Taber JM, McQueen A, Housten AJ, Studts JL, Scherer LD. Translating cancer risk prediction models into personalized cancer risk assessment tools: stumbling blocks and strategies for success. Cancer Epidemiol Biomarkers Prev 2020 Oct 12;29(12):2389-2394. [doi: 10.1158/1055-9965.epi-20-0861]
47. Brewer N. Goals. In: Communicating Risks and Benefits: An Evidence-Based User's Guide. Silver Spring, MD: U.S. Department of Health And Human Services, Food and Drug Administration; 2011:A-10.
48. Zikmund-Fisher BJ. The right tool is what they need, not what we have: a taxonomy of appropriate levels of precision in patient risk communication. Med Care Res Rev 2013 Feb;70(1 Suppl):37S-49S. [doi: 10.1177/1077558712458541] [Medline: 22955699]
49. Edwards A, Unigwe S, Elywn G, Hood K. Effects of communicating individual risks in screening programmes: Cochrane systematic review. Br Med J 2003 Sep 27;327(7417):703-709. [doi: 10.1136/bmj.327.7417.703]
50. Cancer Prevention Overview (PDQ®) - Patient Version. National Cancer Institute. URL: https://www.cancer.gov/about-cancer/causes-prevention/patient-prevention-overview-pdq [accessed 2020-07-07]
51. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. J Natl Cancer Inst 2011 Oct 5;103(19):1436-1443 [FREE Full text] [doi: 10.1093/jnci/djr318] [Medline: 21931068]
52. Net Market Share. 2020. URL: http://netmarketshare.com [accessed 2021-01-05]
53. Ulin P, Robinson E, Tolley E. Qualitative Methods in Public Health: A Field Guide for Applied Research. San Francisco: Jossey-Bass; 2004.
54. Cortez S, Milbrandt M, Kaphingst K, James A, Colditz G. The readability of online breast cancer risk assessment tools. Breast Cancer Res Treat 2015 Nov;154(1):191-199 [FREE Full text] [doi: 10.1007/s10549-015-3601-2] [Medline: 26475705]
55. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. Adv Exp Soc Psychol 1996:123-205. [doi: 10.1016/s0065-2601(08)60214-2]
56. Centers for Disease Control. Impact of a national tobacco education campaign on weekly numbers of quitline calls and website visitors--United States, March 4-June 23, 2013. MMWR Morb Mortal Wkly Rep 2013 Sep 20;62(37):763-767 [FREE Full text] [Medline: 24048152]
57. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Med Care Res Rev 2007 Apr;64(2):169-190. [doi: 10.1177/10775587070640020301] [Medline: 17406019]
58. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. A demonstration of 'less can be more' in risk graphics. Med Decis Making 2010;30(6):661-671 [FREE Full text] [doi: 10.1177/0272989X10364244] [Medline: 20375419]

Abbreviations

HINTS: Health Information National Trends Survey
NCI: National Cancer Institute

Edited by G Eysenbach; submitted 07.08.20; peer-reviewed by M Falahee, C Basch; comments to author 18.11.20; revised version received 19.11.20; accepted 19.12.20; published 25.01.21.

Please cite as:
Waters EA, Foust JL, Scherer LD, McQueen A, Taber JM
Adherence of Internet-Based Cancer Risk Assessment Tools to Best Practices in Risk Communication: Content Analysis
J Med Internet Res 2021;23(1):e23318
URL: http://www.jmir.org/2021/1/e23318/
PMID:33492238

Original Paper

Patient Perspectives on Health Data Privacy and Implications for Adverse Drug Event Documentation and Communication: Qualitative Study

Serena S Small[1,2*], MA; Corinne M Hohl[1,2,3*], MHSc, FRCPC, MD; Ellen Balka[1,4*], PhD

[1]Centre for Clinical Epidemiology & Evaluation, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada
[2]Department of Emergency Medicine, University of British Columbia, Vancouver, BC, Canada
[3]Emergency Department, Vancouver General Hospital, Vancouver, BC, Canada
[4]School of Communication, Simon Fraser University, Burnaby, BC, Canada
[*]all authors contributed equally

Corresponding Author:
Serena S Small, MA
Centre for Clinical Epidemiology & Evaluation
Vancouver Coastal Health Research Institute
710-828 W 10th Ave
Vancouver, BC, V5Z1M9
Canada
Phone: 1 6048754111 ext 55219
Email: serena.small@ubc.ca

Abstract

Background: Adverse drug events are unintended and harmful effects of medication use. Using existing information and communication technologies (ICTs) to increase information sharing about adverse drug events may improve patient care but can introduce concerns about data privacy.

Objective: This study aims to examine the views of patients and their caregivers about data protection when using ICTs to communicate adverse drug event information to improve patient safety.

Methods: We conducted an exploratory qualitative study. A total of 4 focus groups were held among patients who had experienced or were at risk of experiencing an adverse drug event, their family members, and their caregivers. We recruited participants through multiple avenues and iteratively analyzed the data using situational analysis.

Results: Of the 47 participants recruited, 28 attended our focus groups. We identified 3 primary themes. First, participants felt that improved information sharing about adverse drug events within their circle of care would likely improve care. Second, participants were concerned about data handling and inappropriate access but believed that the benefits of information sharing outweighed the risks of privacy breaches. Finally, participants were more concerned about data privacy in the context of stigmatized health conditions.

Conclusions: Current conditions for maintaining health data privacy are consistent with participants' preferences, despite the fact that health data are susceptible to breaches and mismanagement. Information sharing that increases patient safety may justify potential privacy risks. Greater attention to patient concerns and the effect of social and contextual concerns in the design and implementation of health information technologies may increase patient confidence in the privacy of their information.

(J Med Internet Res 2021;23(1):e21452) doi:10.2196/21452

KEYWORDS

health information technology; adverse drug events; privacy of patient data

Introduction

Background

Adverse drug events are unintended and harmful outcomes of medication use and a leading cause of emergency department visits and unplanned hospital admissions [1-4]. More than 30% of patients presenting to hospitals with adverse drug events are affected by repeat events that occur because care providers unintentionally re-expose patients to medications that previously caused harm [5]. A lack of effective automated processes to communicate adverse drug event information between health providers and across locations of care contributes to the recurrence of these events.

Poor communication about adverse drug events reflects broader fragmentation and siloed information in health. Recent initiatives at the provincial and federal levels of the government in Canada aim to address these communication gaps. In Ontario, the Digital First for Health Strategy intends to increase the availability of patient records for frontline clinicians and reduce barriers to integration [6]. British Columbia's Digital Health Strategy seeks to modernize the health system through integration, improved care delivery, and data accessibility for clinicians and patients [7]. This includes increasing access to clinical information through end-to-end medication management using existing clinical information systems [8]. Enabling PharmaNet, British Columbia's medication dispensing database, to receive and transmit adverse drug event information may support such an undertaking.

Objectives

Enhanced communication about adverse drug events may improve patient safety but can also reveal sensitive diagnoses to a broader range of clinicians than those currently aware of them. This could introduce privacy concerns for patients, particularly among those living with stigmatized illnesses. We examined patients' perceptions about the need to share information about adverse drug events to optimize patient safety while maintaining data privacy.

Methods

Study Design

We conducted a qualitative study to explore patients' perceptions of information privacy and sharing in the context of developing software to facilitate adverse drug event documentation and communication. Research ethics boards of the University of British Columbia and Simon Fraser University reviewed and approved the protocol. All participants provided written informed consent.

Study Setting and Sample

Our target study population included adults (≥19 years) who had lived experience with or were at risk of an adverse drug event and family members and caregivers of patients who had experienced an adverse drug event. All participants lived in the Vancouver area or Whistler, British Columbia, between September 2016 and May 2017. We excluded patients who were living in long-term care facilities, did not manage their own medications, were receiving palliative care, were from out of province, or did not speak English. If we approached a patient who was excluded based on the above criteria, we attempted to recruit a family member or caregiver for participation, if they were present during recruitment.

Recruitment

We used multiple sampling strategies to recruit patients who had lived experience with an adverse drug event or who were at risk of an adverse drug event because of their age (≥65 years) or exposure to polypharmacy and the family members and caregivers of patients who had experienced an adverse drug event. We recruited in person, through posters, and through web-based advertisements.

From September to November 2016, we recruited those who were experiencing or who were at risk of experiencing an adverse drug event. Emergency department pharmacists recruited a convenience sample of patients presenting to the emergency department of Vancouver General Hospital, a tertiary care hospital in Vancouver, Canada, when completing medication reviews. We placed posters in high-traffic areas in the emergency department and at a hospital-based research center to encourage patients and their family members to contact the research team if they were interested in participating. We also sought to recruit members of the general population by posting web-based advertisements on Kijiji and Craigslist and by snowball sampling from personal connections. In recruitment advertisements, we stated that we were seeking patient opinions on having information about their medication-related problems shared among care providers. Our intent was to reach a broader range of individuals who met our target sample criteria, including those that may not have had direct contact with the acute care setting at the time of the focus groups.

From February to May 2017, we recruited individuals with stigmatized illnesses (HIV and/or substance use disorder) by placing posters at a clinic that provides care to HIV-positive women. We hypothesized that these patients may have specific privacy concerns and also because HIV medications are currently not documented in the provincial medication dispensing database, PharmaNet, in part because of privacy concerns at the time of PharmaNet's implementation [9].

Among those recruited through snowball sampling from personal connections, there was an established relationship between the researchers and participants before the study. For all others, beyond contact for recruitment purposes, there was no prior established relationship.

Data Collection

The focus groups followed a semistructured discussion guide developed collaboratively by the research team to address themes relevant to adverse drug event information sharing (Multimedia Appendix 1). The principal investigator (EB) and a research assistant (SS) with expertise in qualitative research created the first draft. Other members of the research team then revised and edited the discussion guide to offer different disciplinary perspectives.

Key themes addressed in the discussion guide were experiences with adverse drug events, knowledge of information-sharing practices, and attitudes about data privacy and privacy policy. We allowed participants to engage in open dialog and ask questions beyond the discussion guide and identify and discuss new concepts that we had not considered.

We held focus groups in a research office at the Vancouver General Hospital. The principal investigator (EB), a female social scientist with extensive experience in qualitative methods, led the focus groups, and a research assistant (SS) attended to take notes. At the beginning of each group, we introduced the researchers present and provided a brief definition of adverse drug events, including examples, providing rationale for the groups. Recruitment materials informed prospective participants that we sought to gather opinions to guide the development of a system to support sharing of information about adverse drug events among health care providers. We reiterated this at the time of the focus groups.

We gathered additional information about the participants, including demographic information, using a short debriefing survey at the conclusion of each focus group (Multimedia Appendix 2). We audio recorded the focus groups, which were then transcribed by a research assistant (SS).

Data Analysis

We coded and analyzed transcriptions using *NVivo 11* qualitative data analysis software (QSR International, version 11, 2015). We created a provisional coding frame to reflect the thematic structure and discussion guide questions. The structure of the coding frame organized participant comments conceptually along the following themes: data privacy, information sharing, awareness of privacy policy, policy preferences, experience with adverse drug events, and recommendations. We (SS and EB) iteratively coded and analyzed the data using situational analysis, a theoretical and methodological approach that examines contextual, relational, and discursive elements in the data through the concurrent creation of memos and mapping exercises [10].

Results

Focus Groups and Participant Characteristics

Of the 47 participants we recruited, 28 attended a focus group. Each focus group had 5 to 8 participants (Table 1). A total of 20 participants (20/28, 71%) were in groups A, B, and C. Of these participants, 65% (13/20) were aged above 65 years and at risk of an adverse drug event, 25% (5/20) were from the general population, and 10% (2/20) were caregivers or family members of patients with adverse drug events. Group D consisted of 8 women (8/28, 29%) recruited from a clinic serving HIV-positive women.

Table 1. Focus group composition (N=28).

Group ID	Participants, n (%)
A	8 (29)
B	7 (25)
C	5 (18)
D	8 (29)

Each focus group lasted for 120 min. Most participants (25/28, 89%) completed the debriefing survey (Table 2). Most participants were female and aged above 51 years. Many participants had lived experiences with an adverse drug event, knew someone else who had, or both. Most participants had completed at least some postsecondary education.

Table 2. Participants’ characteristics (N=28).

Variable	Participants, n (%)
Gender	
Male	6 (21)
Female	19 (68)
No response	3 (11)
Age (years)	
<20	0 (0)
20-35	2 (7)
36-50	5 (18)
51-65	10 (36)
>66	8 (29)
No response	3 (11)
Experience with adverse drug events	
Yes, have lived experienced with an adverse drug event	3 (11)
Know someone who has experience with an adverse drug event	7 (25)
Both lived experience and know others who have experienced an adverse drug event	6 (21)
No, have not experienced an adverse drug event	6 (21)
Unsure	3 (11)
No response	3 (11)
Highest level of education	
Some high school	3 (11)
Completed high school	3 (11)
Some postsecondary	7 (25)
Completed college or university	7 (25)
Some graduate school	2 (7)
Master’s degree	1 (4)
Doctoral degree	1 (4)
No response	4 (14)

Primary Themes

We identified 3 primary themes about information sharing and privacy in the context of adverse drug event communication. Participant quotes to support each theme are presented in the corresponding textboxes, which are representative of the findings for each theme.

Participants Believed Enhanced Information Sharing Among Clinicians Would Improve Care

Experiences With Informational Discontinuity of Care

Many participants described experiences with fragmented information sharing (Textbox 1). In some cases, participants experienced negative outcomes as a result of poor information sharing. Participant 1, for example, described how their father-in-law’s experience with poor communication of an adverse drug event affected his long-term health, resulted in unnecessary costs, and emotionally affected the patient and his family. Several participants noted that in the absence of effective information-sharing processes, they took responsibility for information sharing themselves.

Textbox 1. Participant quotes about experiences with informational discontinuity of care.

- “I don’t think there’s a lot of communication between the doctors. Say you have a GP, you have a rheumatologist, you have an HIV specialist – you tell one doctor one thing and they say, ‘oh I didn’t know that, when did that start?’…So you gotta follow up yourself because he can’t…they get so busy, or they forget, or they don’t care or whatever.” [Participant 25, group D]
- “I have been to [the] emergency department with an adverse reaction twice, and it’s very busy. The doctors who see you don’t have time to write your discharge report up in time for you to take a copy away with you…I would like to have further information and would like to be able to pass it on to my GP.” [Participant 11, group B]
- “I think in my father-in-law’s case, [the lack of information sharing] cost our system more money…there were more doctors involved…there was more angst involved…my father-in-law’s condition plummeted. And there were more people involved, there was more testing done, ambulance was called ten times.” [Participant 1, group A]
- “I don’t think it’s very well shared. If it is, it’s pretty piecemeal.” [Participant 3, group A]
- “I’ve just always heard that adverse effects are supposed to be reported…but I never had any confidence that they were.” [Participant 8, group A]
- “I have learned that in many cases, the people to whom I go for one medical event or another don’t always share the information.” [Participant 12, group B]
- “My family doctor gets everything, but I particularly have to make a point of asking for copies to be sent to a couple of my specialist physicians.” [Participant 11, group B]

Benefits of Better Communication

Participants felt that better communication between providers in their circle of care would have a positive effect on health outcomes and could improve their experience with the health system, including improved disease and medication management (Textbox 2). They suggested that communication would lessen the recall burden for patients and their families and that broader information sharing would support clinical decision making, especially in situations in which a patient would be unable to communicate or recall the required information.

Textbox 2. Participant quotes describing the benefits of better communication.

- “And then…someone who is elderly, who may have dementia, who doesn’t have someone advocating for [them] – that information needs to be shared so that somebody can make sure that they’re making good decisions around their health care and prescription medications.” [Participant 1, group A]
- “[My father-in-law’s] health would have been maintained at a higher level for a longer period of time, had the information been shared more regularly.” [Participant 1, group A]
- “Yes, the more [my care providers] know [about my] medications…the better they’re taking care of me.” [Participant 27, group D]
- “When my mother was admitted and all they need is her care card number…all the information [is] there already, so it’s a lot easier for us.” [Participant 19, group C]
- “You’d think that the more information that your caregivers have…the better off you’re going to be if you have a problem or if you’re unconscious or whatever.” [Participant 20, group C]
- “I think the more we share the information, it’s a huge financial benefit…both the emotional and financial side.” [Participant 7, group A]
- “…We have the language barriers, we have culture barriers, we have all of those things to deal with and it makes things very, very difficult, so another reason for having this information [available] to so many people.” [Participant 3, group A]
- “My view is that I am less concerned about privacy, and more concerned about people [caring] for me having the information that they need.” [Participant 8, group A]

Most Participants Preferred Electronic Information Sharing

Many participants supported the use of health information technologies to share adverse drug events and medication information (Textbox 3). Participants viewed electronic communication as quick, easy, and environmentally sustainable while also reducing the risk of lost or misplaced files. Although recognizing these advantages, participants were concerned about data security threats (eg, hackers) and system failure (eg, because of an earthquake). As a result, participants felt that clinicians should not rely exclusively on electronic information sharing and storage. Participants suggested backups to electronic information sharing, including telephone-based communications between clinicians, and electronic or paper-based backups.

Textbox 3. Participant quotes about electronic information sharing.

- “I personally would want it electronically, just because with the technology age nowadays it’s…the easiest method and the quickest method to transfer information…also it’s more eco-friendly than having all these pieces of paper that might get lost…” [Participant 18, group C]
- “I think I’d want it electronic because it is simpler, and mounds of more paper aren’t necessary…or needed.” [Participant 16, group C]
- “All I’ve gotta say is: is there an app for that?” [Participant 3, group A]
- “I don’t think we should lose the ability to ever…pick up the phone, because there are quick emergency situations that can save a life…but generally I think that electronically is the most practical.” [Participant 7, group A]
- “I’m in favour [of] electronically, but with ensured back up because you could lose everything.” [Participant 8, group A]
- “Digital is my first choice.” [Participant 14, group B]
- “There’s been breaches with confidential medical files when it comes to computers.” [Participant 21, group D]
- “I would want it electronically and [on] paper because…we live in an earthquake zone.” [Participant 17, group C]

Participants Were Concerned With Data Handling and Inappropriate Access

Participants Believed Professional Roles Should Determine Access Permissions

Participants focused on the different professional groups that would access their health information rather than the information systems that would mediate information sharing (Textbox 4).

Role-based access was a recurrent theme, and participants discussed whether access to information was pertinent to every clinical role. For example, group A agreed that pharmacists needed access to patient information but debated whether pharmacy assistants also did. Similarly, participants in groups A and D questioned whether care providers in long-term care facilities (eg, care aides) or allied health professionals (eg, physiotherapists) required full access to medical information or if they had adequate training to manage confidentiality.

Textbox 4. Participant quotes about role-based access.

- “I would say it would be ok for [my information to be shared] as long as it’s…the doctor... Say…someone on a team…like maybe a social worker of something, they might be valuable on the team, but the medication part would have absolutely nothing to do with them. So, they shouldn’t be having access to that information because they can’t do anything about it.” [Participant 21, group D]
- “When it comes to doctors and nurses…they share. But if…the definition of ‘care team’ is broader than that, then I would need to know who they were and what they were doing [with my information.]” [Participant 17, group C]
- “[Care aides are] usually wonderful people, but they don’t have the information…on how to deal with the ethics of private information. This is my experience with my mother. But yeah, it would depend who it was. I mean my physio doesn’t need to know, right?” [Participant 17, group C]
- “It’s more of a question of what is their education, what is their guidelines…Like a pharmacist…keeps everything confidential. Does the assistant?” [Participant 7, group A]
- “Maybe…the key is who gets the information [is] anybody that has to do with the prescription.” [Participant 3, group A]
- “You run into a whole hornet’s nest when you’re talking about other people getting that information, like for example insurance companies.” [Participant 12, group B]
- “[The doctors] always have [medical] students in their office, right? They come in with the doctors…So the [medical student has] your information, they’ve got everything in that conversation…And where does it go from there?” [Participant 25, group D]
- “I was just going to say that all the medical and allied health, secondary health professions, have confidentiality and privacy as a really major, serious part of their curriculum…It’s as secure as it can be given people.” [Participant 8, group A]

Participants Perceived Internal and External Threats to Their Information in Medical Facilities

Participants believed that data stored electronically in medical facilities were unlikely to be secure (Textbox 5). When asked whether they were aware of any breaches of medical information, a small number of participants said yes. At least one breach was mentioned in each group. Several said that they had heard of both clinical and administrative staff mishandling data, including improper disposal, private conversations in public spaces, and inappropriate access of records. Participants had also heard of external threats, including hacking and breaches of Canadian data by American companies; however, none had been firsthand victims of data breaches in medical facilities. Despite this awareness, most agreed that hearing about breaches did not affect their willingness to share their health data. One participant summarized this sentiment by stating that they thought the benefits of information sharing exceeded the risks of privacy breaches.

Textbox 5. Participant quotes about data security.

- “I think if someone really wanted to get [my medical information] they could get it…People within the office [of] the hospital, if they wanted to get access to it, I think they could get it. Even if they don’t have…official approvals. And, data hacking is advancing.” [Participant 17, group C]
- “I never thought of it as being that confidential…It’s on a computer and in your pharmacy, and there’s a zillion people [that] have access to it. I just accept the system the way it is.” [Participant 20, group C]
- “[Information is] only as secure as the individuals handling it.” [Participant 3, group A]
- “I don’t think pharmacies are very secure.” [Participant 15, group B]
- “There’s a lot of hackers out there that can get access to [your information].” [Participant 25, group D]
- “There’s faxes that have…gone to the wrong fax number, so there’s a breach of confidentiality there.” [Participant 21, group D]
- “And where the breaches occur are chatting down the hallway, nurses chatting in the elevator, people in the cafeteria…” [Participant 8, group A]
- “I’ve heard of…an operator throwing some CDs or…storage device in the garbage and then somebody went in the garbage and pulled it out, and there’s half a million records on there.” [Participant 3, group B]
- “I mean, there have been serious privacy breaches on record with the provincial government specifically.” [Participant 15, group B]
- “A whole stack of personal information and somebody just dumped it out in the back lane somewhere.” [Participant 14, group B]

Privacy Concerns Are Amplified When Considering Stigma and Potential Discrimination

Discussions about stigma and discrimination around health data emerged in most groups (Textbox 6). Many felt that sharing information about stigmatized illnesses, which could occur if an adverse drug event to HIV medication was recorded, should occur only within a patient’s circle of care, which is consistent with current data privacy standards. Among participants in group D (who live with stigmatized illnesses), concerns about the effects of stigma and discrimination were amplified. Previous experience with the health system that had reduced complex lived experiences with negative labels colored this group’s perception of the system. There was a relationship between trust, willingness to share health information, and stigma. One participant, for example, commented that they might withhold medical information if they did not trust their care provider or had concerns about where their information was going and who could access it. Although we sought to understand information-sharing preferences in clinical settings, several participants mentioned unprompted that their sensitive medical information should not be shared with colleagues or employers. Participants suggested ways to reduce stigma, including educating clinicians and providing information for patients during care encounters.

Textbox 6. Participant quotes about stigma and discrimination.

- “When [a care provider] comes in and [is] talking about someone’s health, you just don’t outright say ‘hey how’d you get that?’ That’s really disrespectful to a person whether they have diabetes, cancer, HIV, or Hep C, or whatever their medical situation is.” [Participant 21, group D]
- “So, depending on who’s accessing that information…that’s where stigma, discrimination comes in, because [the patient] could…end up being judged from the medication they’re on because [the care provider] knows what those medications are used for.” [Participant 21, group D]
- “I remember all the stigma around cancer when I was a child. It was like the ‘c-word’. You didn’t even call it cancer. And it’s great to see the shift now…It would be lovely if we could get that way with stuff like mental health, Hep C, HIV. You know, there’s definitely more awareness out there but unfortunately there’s…the ignorance.” [Participant 21, group D]
- “I’ll tell you right now, the honest truth about the Downtown Eastside is [care providers] don’t care about you. You’re just a number, you’re just an addict, you’re just a prostitute, you’re just a drunk…You’re not a human being…You’re just shuffled through, seen by whoever’s there [at the clinic] …They don’t [have] your files.” [Participant 28, group D]
- “There was the case of the woman refused entry into the US because she was on anti-depressants.” [Participant 17, group C]
- “I have a mental illness…if I’m a danger to myself or to others, and I’m not taking my meds…then yes, my diagnosis along with my meds need to be passed on to somebody…if I’m stable, then just my medication [information should be shared].” [Participant 4, group A]
- “You know, [doctors] want you to tell them everything…but I don’t want to tell you [doctors] this part, because I don’t trust you guys.” [Participant 25, group D]
- “I just feel…the students…or the…residents, they need to be better educated on…bedside manners.” [Participant 21, group D]
- “I do see more often that…the confidentiality blurbs are out there more often when you’re signing things, [saying] this is how we protect information. That never used to be out there, so I think there is more awareness out there, but I think it needs to continue, like even…ramp it up. And not [allow] people [to] get…complacent about it.” [Participant 21, group D]

Discussion

Principal Findings

We explored patients' perceptions about information privacy and sharing in the context of developing a health information technology that will enable electronic documentation and automated communication of adverse drug event information between providers and across care settings. Most participants supported improved information sharing about adverse drug events, expected technologies, and clinicians to protect their privacy and understood that a lack of information sharing could pose a greater risk to their safety than potential threats to privacy. In the following sections, we explore how existing organizational and institutional measures to protect data privacy are consistent with participant expectations.

Privacy legislation provides a framework for data management. In British Columbia, the Freedom of Information and Protection of Privacy Act (FIPPA) governs the collection, use, and dissemination of personal information by public entities, including health authorities and hospitals. Under FIPPA, public entities must enact and enforce security measures to prevent unauthorized collection, use, access, disclosure, and disposal of personal information. Public entities such as hospitals and clinics must notify individuals that their information is being collected and used but do not require patient consent to share patient information with other members of the care team [11]. Our study participants' privacy preferences were consistent with these legislative requirements: participants were favorable toward information sharing among clinicians but were wary of giving access to those outside their circle of care (eg, an insurance company) or among those within the circle of care for whom information about an adverse drug event is irrelevant to their role (eg, a physiotherapist). A system that supports adverse drug event data sharing among a patient's circle of care must minimize the barriers to effective communication and should not require additional patient consent.

Improvement in communication about adverse drug events may be achieved by leveraging existing health information technologies, such as PharmaNet. PharmaNet employs numerous data management safeguards, including physical security (eg, limited access to equipment), operating system security (eg, user access keys), network security (eg, firewall), and screen security (eg, only certain items viewable based on each user's security profile) [12]. It also adheres to the principles of role-based access, wherein different user groups have different access permissions, which was strongly preferred among participants. In addition, individuals can find out when their record is accessed and those with further privacy concerns can add a password to their PharmaNet profile, allowing them to determine who can or cannot view their profile. This is valuable for individuals living with stigmatized illnesses.

Despite the privacy measures implemented in health information technologies, security under real-world conditions is more volatile. Breaches of health data have been a recurrent focus of media attention and critique, including incidents involving PharmaNet. In 2014, for example, approximately 1600 profiles were compromised by an unknown, unauthorized individual using a doctor's account [13]. In 2017, more than 20,000 profiles were breached [14]. These breaches exemplify our participants' concerns and demonstrate the challenge of managing the risks associated with privacy breaches while ensuring that data are accessible in the interest of patient care. Following these events and other privacy concerns, the provincial Ministry of Health introduced a new project to support user management for PharmaNet, which will streamline access approval when implemented [15].

In addition to legislative frameworks, concerns about handling sensitive health information can be addressed in the design of systems by implementing role-based access functionality, building complex password requirements, and regularly auditing use and users. Participants' concerns regarding threats to data security among staff in medical facilities can be addressed through other nontechnical approaches. Implementation should incorporate education that addresses safe information handling, including proper methods of sharing data and disposing of paper records, and strategies for maintaining the security of log-in credentials. These measures may increase clinicians' ability to maintain the security of information in their custody while increasing patients' confidence in the privacy of their information and in the efficacy of information sharing in health.

Limitations

Sample composition is the primary limitation of this study. Participant self-selection and recruitment from an urban area may have introduced selection bias. More women participated than men. Participants in rural regions, men, and those with other health trajectories or access points within the health system may have different experiences in the health sector that are not reflected. As such, our findings may not translate to other regions, populations, and health conditions. In addition, we did not screen participants who volunteered via classified websites (n=5) to determine whether they met the defined sample criteria (ie, at risk of or experienced an adverse drug event or a family member or caregiver). Therefore, we cannot verify whether all responses are representative of these sample criteria.

Conclusions

Participants were generally supportive of enhanced informational continuity of care about adverse drug events to facilitate care delivery. The belief that enhanced information sharing would improve care and that a lack of information sharing poses safety risks indicates patient support for broader use of information and communication technologies (ICTs) in health. Privacy considerations were important to participants but largely in the context of the human actors handling the data rather than the electronic systems that mediate information transfer. Fears about stigma and discrimination were prominent drivers, particularly among patients who had experienced stigmatization. Our findings suggest the need to consider the ways that social and contextual factors (eg, living with a stigmatized illness) that affect patient privacy can be addressed at both the human and technical levels in the design and implementation of ICTs in health.

Acknowledgments

This research was sponsored by the Canadian Institutes of Health Research eHealth Innovations Partnership Program (no. 348320) and the Michael Smith Foundation for Health Research (no. 16724).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Discussion guide.
[DOCX File , 21 KB - jmir_v23i1e21452_app1.docx]

Multimedia Appendix 2
Follow-up survey.
[DOCX File , 20 KB - jmir_v23i1e21452_app2.docx]

References

1. Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Ann Intern Med 2004 May 18;140(10):795-801. [doi: 10.7326/0003-4819-140-10-200405180-00009] [Medline: 15148066]
2. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998 Apr 15;279(15):1200-1205. [doi: 10.1001/jama.279.15.1200] [Medline: 9555760]
3. Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. N Engl J Med 2011 Nov 24;365(21):2002-2012. [doi: 10.1056/NEJMsa1103053] [Medline: 22111719]
4. Zed PJ, Abu-Laban RB, Balen RM, Loewen PS, Hohl CM, Brubacher JR, et al. Incidence, severity and preventability of medication-related visits to the emergency department: a prospective study. CMAJ 2008 Jun 03;178(12):1563-1569 [FREE Full text] [doi: 10.1503/cmaj.071594] [Medline: 18519904]
5. Peddie D, Small SS, Badke K, Wickham ME, Bailey C, Chruscicki A, et al. Designing an adverse drug event reporting system to prevent unintentional reexposures to harmful drugs: study protocol for a multiple methods design. JMIR Res Protoc 2016 Aug 18;5(3):e169 [FREE Full text] [doi: 10.2196/resprot.5967] [Medline: 27538362]
6. Ontario Ministry of Health. Ontario expanding digital and virtual health care: giving patients more options part of province's plan to end hallway health care. Government of Ontario Newsroom. 2019 Nov 13. URL: https://news.ontario.ca/en/release/54594/ontario-expanding-digital-and-virtual-health-care [accessed 2021-01-04]
7. Patients as partners: 2019 community conversations. Government of British Columbia. 2019 Mar 04. URL: https://www2.gov.bc.ca/assets/gov/health/about-bc-s-health-care-system/heath-care-partners/patients-as-partners/2019_community_conversations_summary_report_-_final.pdf [accessed 2020-12-23]
8. BC Health Information Management Professionals Society (BCHIMPS) 2019 Annual Spring Education Symposium. Province of BC Digital Health Strategy (DRAFT): Transforming the health system so that all British Columbians can achieve optimal health & wellness. 2019 Mar 01. URL: https://www.bchimps.org/resources/Documents/2019%20Spring%20Conference/HLTH%20ADMs_Barclay_Wright_1Mar2019.pdf [accessed 2020-12-24]
9. Kendall PRW. Provincial health officer's report on HIV reportability. British Columbia Ministry of Health Planning. 2002 Feb. URL: https://www.health.gov.bc.ca/library/publications/year/2002/hivreportability.pdf [accessed 2020-12-22]
10. Clarke AE. Situational Analysis: Grounded Theory after the Postmodern Turn. In: Sage Research Methods. Thousand Oaks, CA: Sage Publications, Inc; 2005.
11. Guide to access and privacy protection under FIPPA. Office of the Information and Privacy Commissioner of BC. 2015 Oct. URL: https://www.oipc.bc.ca/guidance-documents/1466 [accessed 2020-12-23]
12. PharmaCare Policy Manual. Government of British Columbia. 2012 Mar. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/pharmacare/pharmacare-publications/pharmacare-policy-manual-2012#section-9 [accessed 2020-12-24]
13. BC government warns 1,600 people about breach of personal information. The Canadian Press. 2014 Jul 11. URL: https://infotel.ca/newsitem/bc-government-warns-1600-people-of-pharmanet-privacy-breach/it11473 [accessed 2021-01-04]
14. Police announce arrest following alleged breach of B.C.’s PharmaNet system. Canadian Press. 2017 Apr 03. URL: https://www.vicnews.com/news/police-announce-arrest-following-alleged-breach-of-b-c-s-pharmanet-system-2/ [accessed 2020-12-14]

15. PRIME: PharmaNet Revisions for Information Management Enhancements. Government of British Columbia. -. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/pharmacare/pharmanet-bc-s-drug-information-network/prime [accessed 2021-01-03]

Abbreviations

FIPPA: Freedom of Information and Protection of Privacy Act
ICT: information and communication technology

Edited by G Eysenbach; submitted 15.06.20; peer-reviewed by A Thompson, K Goodman; comments to author 21.07.20; revised version received 29.07.20; accepted 28.10.20; published 20.01.21.

Please cite as:
Small SS, Hohl CM, Balka E
Patient Perspectives on Health Data Privacy and Implications for Adverse Drug Event Documentation and Communication: Qualitative Study
J Med Internet Res 2021;23(1):e21452
URL: http://www.jmir.org/2021/1/e21452/
doi:10.2196/21452
PMID:33470936

Original Paper

Improving Uptake of a National Web-Based Psychoeducational Workshop for Informal Caregivers of Veterans: Mixed Methods Implementation Evaluation

Veronica Yank[1], MD; Randall C Gale[2], DrPH; Andrea Nevedal[2], PhD; Leonore Okwara[3], MPH; Christopher J Koenig[4], PhD; Ranak B Trivedi[2,5], PhD; Nancy J Dupke[6], MSW; Margaret Kabat[6], LCSW-C, CCM; Steven M Asch[2,7], MD, MPH

[1]Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, United States
[2]Center for Innovation to Implementation, VA Palo Alto Health Care System, Palo Alto, CA, United States
[3]Department of Behavioral and Community Health, University of Maryland, Baltimore, MD, United States
[4]Department of Communication Studies, San Francisco State University, San Francisco, CA, United States
[5]Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA, United States
[6]VA Caregivers Support Program, Department of Veterans Affairs, Washington, DC, United States
[7]Division of Primary Care and Population Health, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Veronica Yank, MD
Division of General Internal Medicine
Department of Medicine
University of California San Francisco
1545 Divisadero Street, Suite 322
San Francisco, CA, 94143-0320
United States
Phone: 1 415 476 9654
Email: veronica.yank@ucsf.edu

Abstract

Background: Although web-based psychoeducational programs may be an efficient, accessible, and scalable option for improving participant well-being, they seldom are sustained beyond trial publication. Implementation evaluations may help optimize program uptake, but few are performed. When the US Department of Veterans Affairs (VA) launched the web-based psychoeducational workshop Building Better Caregivers (BBC) for informal caregivers of veterans nationwide in 2013, the workshop did not enroll as many caregivers as anticipated.

Objective: This study aims to identify the strengths and weaknesses of initial implementation, strategies likely to improve workshop uptake, whether the VA adopted these strategies, and whether workshop enrollment changed.

Methods: We used mixed methods and the Promoting Action on Research Implementation in Health Services (PARIHS) implementation evaluation framework. In stage 1, we conducted semistructured interviews with caregivers, local staff, and regional and national VA leaders and surveys with caregivers and staff. We collected and analyzed survey and interview data concurrently and integrated the results to identify implementation strengths and weaknesses, and strategies likely to improve workshop uptake. In stage 2, we reinterviewed national leaders to determine whether the VA adopted recommended strategies and used national data to determine whether workshop enrollment changed over time.

Results: A total of 54 caregivers (n=32, 59%), staff (n=13, 24%), and regional (n=5, 9%) and national (n=4, 7%) leaders were interviewed. We received survey responses from 72% (23/32) of caregivers and 77% (10/13) of local staff. In stage 1, survey and interview results were consistent across multiple PARIHS constructs. Although participants from low-enrollment centers reported fewer implementation strengths and more weaknesses, qualitative themes were consistent across high- and low-enrollment centers, and across caregiver, staff, and leadership respondent groups. Identified strengths included belief in a positive workshop impact and the use of some successful outreach approaches. Implementation weaknesses included missed opportunities to improve outreach and to better support local staff. From these, we identified and recommended new and enhanced implementation strategies—increased investment in outreach and marketing capabilities; tailoring outreach strategies to multiple stakeholder groups; use of campaigns that are personal, repeated, and detailed, and have diverse delivery options; recurrent training and mentoring for new staff; and comprehensive data management and reporting capabilities. In stage 2, we determined that the VA had adopted several of these strategies in 2016. In the 3 years before and after adoption, cumulative BBC enrollment increased from 2139 (2013-2015) to 4030 (2016-2018) caregivers.

Conclusions: This study expands the limited implementation science literature on best practices to use when implementing web-based psychoeducational programs. We found that robust outreach and marketing strategies and support for local staff were critical to the implementation success of the BBC workshop. Other health systems may want to deploy these strategies when implementing their web-based programs.

(J Med Internet Res 2021;23(1):e16495) doi:10.2196/16495

KEYWORDS

web-based workshop; implementation science; behavioral intervention technology; informal caregiver; family caregiver; veteran; psychoeducation; self-management; outreach; marketing

Introduction

Background

Evidence-based online psychoeducational programs may offer an efficient, accessible, and scalable option for improving participants' health and well-being [1-9] but only if they succeed in reaching the participants for whom they are designed [10,11]. Unfortunately, few web-based interventions are widely implemented following study completion [3,12]. Programs for informal caregivers offer a case in point [13]. Informal caregivers are family members, friends, or acquaintances (hereafter referred to as *caregivers*) who provide essential support to persons living with major health conditions or disabilities in the community [14,15]. Although web-based psychoeducational programs for caregivers have proliferated, the best strategies for optimizing their implementation remain unknown [7,16-19].

Among caregivers, those providing support to US veterans have some of the highest documented rates of caregiving burden, stress, depression, and neglect of self-care [20,21]. They have expressed a desire for help from the US Department of Veterans Affairs (VA) Health Care System, which is the largest integrated health care system in the United States [20]. Galvanized by a 2010 Act of the US Congress [22], the VA Caregiver Support Program (CSP) commissioned the development of a web-based psychoeducational and self-management workshop, *Building Better Caregivers* (BBC), for these caregivers. In earlier studies, BBC participants experienced reduced stress and depression and increased self-efficacy [23,24]. On the basis of this evidence, the VA adopted the BBC workshop nationally in 2013 at all VA medical centers in the United States and its territories [22]. However, BBC program uptake was lower than expected at many medical centers.

Objectives

We seek to determine why the national web-based BBC workshop did not reach caregivers at some centers and whether uptake could be improved by performing a multistage mixed methods implementation evaluation involving national and regional VA leadership, staff and informal caregivers at local centers, and national VA CSP operations data sets. Specifically, our objectives were organized into 2 stages:

Stage 1: initial evaluation.

- 1a. Identify initial implementation strengths and weaknesses.
- 1b. Identify and recommend implementation strategies likely to improve uptake.

Stage 2: follow-up evaluation.

- 2a. Determine whether new or enhanced implementation strategies were subsequently adopted.
- 2b. Determine whether workshop enrollment changed.

Methods

Study Design

We employed a convergent mixed methods design for both stages of research because the implementation of a web-based caregiver workshop nationally among VA medical centers was a complex health services intervention that required multilevel processes, which we felt would be best assessed with a mixed methods approach [25]. We used the revised Promoting Action on Research Implementation in Health Services (PARIHS) framework as our implementation evaluation framework [26,27]. The PARIHS constructs include (1) evidence (eg, research evidence supporting innovation effectiveness, participant experiences); (2) context (eg, resources necessary to support staff activities, data management and evaluation capabilities); (3) facilitation (eg, skills and attributes of local staff, staff training, and mentoring and assistance from experts); and (4) successful—or unsuccessful—implementation (eg, achievement of desired outcomes of innovation, extent of uptake). These PARIHS constructs were represented throughout our mixed methods process, including in the structure of our sampling frame, survey questions, interview questions, and the analysis and interpretation of both quantitative and qualitative data. The Stanford University Institutional Review Board approved the evaluation (protocol no. 29965), and participants provided verbal consent. Caregivers received US $40 as compensation for study participation. The staff and leaders did not receive compensation.

Stage 1: Identification of Initial Implementation Strengths and Weaknesses and Strategies Likely to Improve Workshop Uptake

In stage 1, we collected and assessed survey and interview data concurrently to enable planned triangulation between the data sets and results (October 2014 to January 2015). The qualitative (interview) approach was the higher priority approach because it enabled the collection of detailed, nuanced feedback on implementation strengths and weaknesses and suggestions for changes not achievable in a survey. However, the quantitative (survey) approach was integral to the implementation evaluation because it enabled anonymized structured feedback on prespecified, discrete implementation elements derived from the PARIHS constructs and subconstructs that might or might not be mentioned during interviews and that we could compare and integrate with emergent themes from the qualitative data. Surveys also generated contextualizing data on respondent characteristics (eg, age, caregiving relationship, care partner health conditions). The quantitative and qualitative analyses were performed by the study team, and the study team, the advisory board, and a national VA CSP leader integrated findings and reached consensus on strategies likely to improve workshop uptake during a half-day virtual retreat (May 2015). The study team communicated findings and recommendations to VA CSP national leaders in a final report and webinar meetings (June to July 2015).

Characteristics of the Workshop

The BBC workshop is an evidence-based, 6-week, interactive psychoeducational and self-management program designed to reduce caregiver stress and depression and increase caregiver skills and self-efficacy [23,24]. Each workshop includes 20-25 participants and is guided by 2 trained cofacilitators. During workshops, caregivers read weekly lessons, perform self-management- and caregiving-related activities to solidify new skills and interact with and receive social support from peers and facilitators. Examples of self-management and caregiving skills include techniques for stress reduction, creating weekly action plans that are realistic and achievable, problem solving when challenges occur, managing difficult care partner behaviors, communicating with family, friends, and health professionals, and planning for the future, among others. The workshop is delivered on a secure online platform developed specifically for the workshop that, at the time, supported use on a computer or tablet device and has since expanded to include smartphone delivery. The workshop is only available in English.

Workshop facilitators are caregivers of persons with chronic conditions. They receive a 5-day training that teaches them facilitation skills and about the workshop (its theoretical basis, structure, and content). Facilitators must demonstrate mastery of the material and adhere to a comprehensive workshop facilitator manual. Master facilitators monitor each workshop for fidelity. Facilitators can lead workshops remotely from anywhere in the United States and are paid to do so.

Implementation Evaluation Context and Advisory Board

The VA adopted the BBC workshop nationally in 2013, making it available at no cost to caregivers registered at any VA medical center in the United States or its territories [22]. The VA CSP leadership and our study team at the VA Center for Innovation to Implementation identified that BBC uptake was lower than expected. Using CSP operations data, we determined that 5.74% (1265/22,022) of eligible caregivers enrolled in the workshop in the first 2 years of the rollout, but that center-specific enrollment rates varied widely—from 0% to 37%—with 10.6% (15/142) of all the centers enrolling no participants. Thus, the VA CSP and our study team partnered to perform the current implementation evaluation. As the study team, we convened an advisory board composed of representatives of key stakeholder groups and experts to advise us at all stages of the evaluation. The advisory board included a veteran caregiver, veteran care partner, expert on informal caregiver research, expert on self-management research, clinical psychologist with expertise in eHealth interventions for veterans and their families, and director of web services for the VA.

Eligibility Criteria and Sample Selection

All 4 national leaders involved in the workshop rollout were invited to participate in the implementation evaluation and all agreed to do so. National leaders were defined as those overseeing or directly managing the BBC implementation at the national level—for example, the VA CSP director, national VA BBC program manager, and BBC project director at the national organization contracted to administer the workshop on behalf of the VA. To identify regional leaders, local staff, and caregivers for participation, we used stratified purposive sampling (Table 1). BBC enrollment rates at all VA medical centers (standardized by catchment size) were calculated to identify high- and low-enrollment centers and the regional networks to which they belonged. At the time, there were 142 medical centers belonging to 21 regional networks. Low-enrollment centers were defined as those in the 10th percentile of referrals during the first 2 years of workshop rollout, and high-enrollment centers were those in the 90th percentile. We identified 5 regional networks that encompassed both low- and high-enrollment centers and invited their CSP regional leaders to participate and all agreed. We then selected 2 regions for local evaluation based on geographical diversity and variation in medical center catchment size: the Northeast region with 2 paired medium-sized centers (1 high-enrollment center and 1 low-enrollment center) and the Gulf Coast with 2 paired large centers (1 high-enrollment center and 1 low-enrollment center). At selected centers, we recruited local staff most involved in the workshop rollout—all CSP social workers who worked as caregiver support coordinators at each center (hereafter referred to as *coordinators*) and other staff whom they identified as involved in the BBC implementation. Among the 14 local staff members approached, 13 (93%) agreed to participate. The staff member who declined had been a substitute coordinator for a short period but had returned to unrelated prior duties by the time we contacted her.

Table 1. Participant sampling frame and subsequent enrollment.

Participant characteristics	Sampling frame			Enrollment
	Nation (United States)	Region	Locality	
Targeted characteristics				
Role	National leaders of BBC[a] rollout (all)	Regional leaders (subset)	Local staff and caregivers (subset)	N/A[b]
Setting	2 national offices	5 regional networks	4 medical centers	N/A
N/A	Both in Washington, DC	N/A	N/A	4 national leaders
N/A	N/A	West	N/A	1 regional leader
N/A	N/A	Midwest	N/A	1 regional leader
N/A	N/A	South	N/A	1 regional leader
N/A	N/A	Gulf Coast	2 large centers[c]	1 regional leader; 9 staff; 16 caregivers
N/A	N/A	Northeast	2 medium centers[c]	1 regional leader; 4 staff; 16 caregivers
Final sample	N/A	N/A	N/A	54 total participants

[a] BBC: Building Better Caregivers.

[b] N/A: not applicable.

[c] Within the regions selected for additional evaluation, we identified and included one high-enrollment medical center and one low-enrollment medical center matched by catchment size—large size in the Gulf Coast and medium size in the Northeast.

We also recruited 8 veteran caregivers per center (32 in total). Caregiver eligibility criteria were being registered with the VA as a veteran caregiver and being 18 years of age or older. Exclusion criteria included having limited English proficiency and inability to speak by phone for a telephone interview (eg, due to hearing loss). We used VA operations data to identify all registered caregivers at each center and cross-referenced these with operations data on workshop referrals and enrollments. Sampling was divided into target groups of (1) caregivers enrolled in the workshop, (2) caregivers aware of the workshop but not enrolled, and (3) caregivers unaware of the workshop. We randomly generated the order of attempted contact with caregivers and performed a maximum of 2 telephone calls per caregiver. We attempted to contact 163 caregivers and, of these, 10 (6.1%) had a nonworking phone number and 101 (62.0%) did not respond to repeated attempts. We spoke with 54 (33.1%) caregivers. Among these, <1% (1/163) was ineligible because of limited English proficiency, 1.8% (3/163) called us back and desired to enroll but could not because we had reached our enrollment targets, 11.0% (18/163) actively refused, and 19.6% (32/163) enrolled. Reasons for refusal included no time, having no one to take care of veteran or other household responsibilities during the interview, veteran having passed away, being 39 weeks pregnant, and wanting to answer questions by mail only. All potential participants (caregivers, staff, and leaders) received information on the study objectives and researcher identities before their agreement to participate.

Survey Approach and Data Reporting

The quantitative evaluation consisted of surveys of caregivers and local staff. Caregivers completed paper-based surveys within a week of interview completion (Multimedia Appendix 1). Questions were asked about their demographic and caregiving characteristics and veteran care partners' demographic and health characteristics. Caregivers who participated in the workshop also indicated whether they would recommend it to others and whether it had improved their quality of life using Likert scale response options 1 to 5 that were anchored by descriptors—strongly disagree, disagree, neither disagree or agree, agree, and strongly agree—and the option of answering do not know/not applicable. The staff survey questions were adapted from the Organizational Readiness to Change Assessment (ORCA), a validated implementation sciences survey structured according to PARIHS constructs (Multimedia Appendix 2) [28]. An additional survey question asked respondents to indicate whether, "At our facility, implementation of the Building Better Caregivers program has been a success," using Likert scale response options 1 to 5 anchored by descriptors—strongly disagree, disagree, neither disagree or agree, agree, and strongly agree—and the option of answering "do not know/not applicable." The ORCA is designed for local staff within organizations implementing evidence-based programs. Therefore, it was completed by local staff only (ie, *not* by regional or national leaders). As survey respondent numbers are small, we report results with descriptive statistics rather than comparative statistics—specifically, the number (%) of respondents who gave the response of interest.

Interview Approach and Data Analysis

The qualitative team consisted of 4 coinvestigators, (RG, CK, AN, and VY) with training and expertise in interview-based data collection and analysis as well as content expertise in the following areas: public health and palliative care (RG), sociology and linguistics of medicine (CK), medical anthropology (AN), and caregiving, primary care, and health services research (VY). All 4 were investigators at the National VA Center for Innovation to Implementation based at the Palo

Alto VA Health Care System. With the exception of VY, who worked with one of the national BBC leaders in a separate study 5 years before this study, the qualitative team did not have prior relationships with the study participants. The qualitative team codeveloped the semistructured interview guides based on PARIHS constructs and pilot tested them with 1 caregiver and 2 staff members at the Palo Alto VA Medical Center before finalizing them (Multimedia Appendix 3).

Data collection consisted of 45-min one-to-one phone-based interviews that were audio-recorded and transcribed by professional transcriptionists and field notes that were recorded by interviewers immediately following interview completion. Transcripts were not returned to the participants for comments or corrections. In caregiver interviews, we asked about caregiving responsibilities, experiences with the VA, use of online technology, and preferences for caregiver support, education, and skills training. Among caregivers who had not enrolled but were aware of the workshop, we asked about outreach or enrollment processes they had experienced, why they had not enrolled, and for any other feedback they chose to share. Among caregivers who had enrolled, we asked about reasons for enrollment and feedback on their outreach, enrollment, and workshop experiences, suggestions for improvements, and whether the workshop had impacted them. In interviews with local staff and regional and national leaders, we collected feedback corresponding to all PARIHS constructs, including asking about their role, training, activities, and observations during the BBC rollout and suggestions for improvements.

The qualitative team codeveloped the codebook. Deductive codes were based on PARIHS constructs, whereas inductive codes were developed using content analysis of transcripts of a subset of 10 interviews and accompanying field notes [29]. Following the development of the initial codebook, 3 coders (VY, RG, and AN) applied a rapid analysis matrix coding form to evaluate an additional 13 interviews, with 2 coders coding each interview (double-coding) [30]. The initial matrix coding form consisted of rows for participants, codes as column headers, and transcript quotes and notes in the pertinent cells displayed in an Excel spreadsheet (Microsoft Corp). The full qualitative evaluation team then met and reviewed the matrices, including codes, quotes, and notes. CK (noncoder) examined quotes and codes to ensure consistency and facilitated discussion of these (eg, of areas of agreement or contention or areas that required clarification), suggested modifications or additions to the codes, and elicited team agreement on the final coding form. Coders applied the final coding form to all interview data, including interviews previously coded at earlier stages. The full qualitative team met again to identify, define, and reach consensus on emergent themes. The team identified themes that addressed strengths and weaknesses of workshop implementation, similarities, and differences in participant feedback from low- and high-enrollment sites and possible strategies for optimizing uptake. Participants did not provide feedback on the findings.

Integration of Survey and Interview Results

Next, we convened a half-day virtual meeting with our study team, advisory board, and a CSP representative using web-based conferencing capabilities. Attendees viewed survey and interview results displayed concurrently by the relevant PARIHS construct. Nonstudy team attendees (advisory board and CSP representative) commented on and confirmed the face validity of the results. All attendees then discussed and interpreted the results and reached a consensus on the integrated study findings.

Identification of Strategies Likely to Improve Workshop Uptake

Building from integrated study findings, at the same meeting, attendees identified broad strategies likely to improve workshop uptake and agreed that these strategies should be recommended to VA BBC national leaders. The study team reported both integrated study findings and recommendations to national leaders in the form of a written report and webinar meetings.

Stage 2: Determination Whether New or Enhanced Implementation Strategies Were Adopted and Whether Workshop Enrollment Changed

In 2018, we performed a follow-up evaluation consisting of interviews and longitudinal workshop enrollment data from the VA CSP operations databases to determine whether new or enhanced implementation strategies had been adopted and whether workshop enrollment changed.

Follow-Up Interviews With National Leaders

In spring 2018, we conducted one-on-one phone interviews with 3 of the original 4 national leaders. The 4th leader was no longer working in the same position and was unavailable for the interview. Our goal was to learn about changes that had been made in response to the initial implementation evaluation. Questions focused on whether existing implementation strategies (from the 2013-2014 timeframe) had been enhanced and new strategies had been adopted (eg, based on stage 1 recommendations) and whether the workshop itself had been maintained. We used the same codebook established from the original 2015 interviews to analyze data from the follow-up interviews.

Longitudinal Data on Workshop Enrollment

Following interviews, we used VA CSP operations data to retrospectively evaluate longitudinal workshop enrollment for the VA as a whole from 2013 to 2018. Data were available by VA fiscal year (FY; October 1 to September 30 of the following year). FY data correspond to the following periods of interest in this study:

1. 2013-2014: BBC workshop rolled out nationally.
2. 2015: study team performed initial implementation evaluation.
3. Late 2015 and early 2016: VA adopted expanded implementation strategies (as determined in follow-up interviews).
4. 2016-2018: follow-up period.

We graphically displayed the longitudinal data for visual inspection to determine the potential impact of expanded implementation strategies on workshop enrollment.

Results

Stage 1: Participants

Thirty-two caregivers, 13 local staff, and 9 regional and national leaders participated in surveys and interviews (Table 2). Approximately half of the caregivers (15/32, 47%) and local staff (7/13, 54%) were from low-enrollment centers. A total of 17 (53%) caregivers supported younger veterans from the post-9/11 era of service. All cared for veterans at high risk for morbidity and mortality based on their health conditions. Most were White, female, above the age of 40 years, and married to their care partners. Their care partner relationships (eg, spouse, child) were similar in distribution to those reported by caregivers enrolled in BBC during the first 2 years of workshop rollout (Multimedia Appendix 4). The chronic conditions of their care partners were also broadly similar to those of caregivers enrolled in BBC (Multimedia Appendix 4). Among the local staff, 69% (9/13) were caregiver support coordinators. Among regional and national leaders, 78% (7/9) were VA based and 2 (22%) were from the national organization responsible for administering the workshops on behalf of the VA.

Table 2. Characteristics of caregivers, local staff, and regional and national leaders.

Participant group and characteristics	Participants, n (%)
Caregivers (n=32)	
Gender	
Female	28 (88)
Male	4 (13)
Age by category (years)	
18-29	4 (13)
30-39	3 (9)
40-49	6 (19)
50-59	6 (19)
60 and older	4 (13)
Not stated	9 (28)
Race/ethnicity	
White	20 (63)
Other[a]	3 (9)
Not stated	9 (28)
Relationship with veteran	
Spouse	26 (81)
Child	3 (9)
Sibling	2 (6)
Parent	1 (3)
Number of chronic conditions for which providing support	
3 or more conditions	26 (81)
1-2 conditions[b]	6 (19)
Experience with workshop	
Enrolled	14 (44)
Aware of it but not enrolled	10 (33)
Not aware	8 (25)
Local staff (n=13)	
Social worker in role of caregiver support coordinator	9 (69)
Other staff[c]	4 (31)
Leaders (n=9)	
VA[d] regional position	5 (56)
VA national position	2 (22)
National organization running workshops	2 (22)

[a] Hispanic/Latino (n=1), Native American/Alaskan Native (n=1), and Other (n=1).

[b] All of the conditions are associated with high morbidity and mortality and include traumatic brain injury, posttraumatic stress disorder, schizophrenia, dementia, and chronic obstructive pulmonary disease.

[c] Other staff include 3 primary care social workers and 1 clinical psychologist.

[d] VA: US Department of Veterans Affairs.

Survey Results

A total of 72% (23/32) of caregivers completed the surveys. They reported high rates of increased stress (21/23, 91%), depression symptoms (19/23, 83%), sleep deprivation (19/23, 83%), neglect of healthy physical activity (15/23, 64%), neglect of healthy eating habits (14/23, 61%), and delay of own health care (11/23, 48%) since becoming a caregiver, which are comparable with national data on veteran caregivers [20]. Among caregivers enrolled in the workshop, 71% (10/14) completed surveys. In response to questions that correspond to the PARIHS evidence construct, 100% (10/10) reported that they would recommend the workshop to others and that their quality of life had improved (8/10, 80%) or may have improved (2/10, 20%) as a result of what they had learned in the workshop.

Among the local staff, 10 (77%) completed surveys: 71% (5/7) from low-enrollment centers and 83% (5/6) from high-enrollment centers. Describing whether workshop implementation at their center had been a success, all 5 staff members from high-enrollment centers agreed, whereas none of the 5 from low-enrollment centers agreed. Nonetheless, most felt that the BBC workshop had been accepted by caregivers and was supported by research and that they had received useful informational materials for caregivers and educational materials for themselves (Table 3). In contrast, a minority reported having a regional mentor who could assist them or having a system for tracking caregivers.

Table 3. Local staff survey results according to Promoting Action on Research Implementation in Health Services constructs.

PARIHS[a] construct	Survey question	Staff in agreement with statement (n=10), n (%)	Linkages to Table 4 theme and whether survey responses are consistent with Table 4 theme[b]
1. Evidence—caregiver experiences	BBC[c] has been well-accepted by VA[d] caregivers	7 (70)	• Theme 2a: Yes[e]
2. Evidence—research evidence	BBC is supported by research evidence	7 (70)	• Theme 2b: Yes
3. Context—staff resources	Informational materials are available for caregivers and staff to raise awareness about the workshop	7 (70)	• Theme 3b: Yes[f] • Theme 4d: No[f]
4. Facilitation—staff skills and attributes	Resources to educate staff about workshop structure and content are available	7 (70)	• Theme 5a: Partial[g]
5. Facilitation—availability of external facilitator to mentor and assist staff	A regional mentor who can help answer questions or solve problems is available	3 (30)	• Theme 5a: Yes
6. Context—information technology capabilities	An approach used at our facility to evaluate and improve implementation includes a system of tracking which caregivers have been referred to BBC	4 (40)	• Theme 5b: Yes

[a] PARIHS: Promoting Action on Research Implementation in Health Services.

[b] Linkages (including areas of consistency and inconsistency) between quantitative and qualitative results were discussed during our process of integrating stage 1 findings, as described in the Methods section.

[c] BBC: Building Better Caregivers.

[d] VA: US Department of Veterans Affairs.

[e] Caregiver responses to survey questions on the PARIHS evidence construct are described in the text and are consistent with these staff responses, as well as with Theme 2a.

[f] Explanation for both a *yes* and a *no* designation: qualitative results in the text and Table 4 indicate that certain outreach materials, particularly those useful for post-9/11 caregivers, were readily available (and used by staff) during workshop rollout, which is consistent with Theme 3b. However, outreach materials, mechanisms, and contacts for pre-9/11 caregivers and others were insufficient (as summarized by Theme 4d), which explains why this survey finding was felt to be inconsistent with Theme 4d.

[g] Explanation for *partial* designation: qualitative results in the text and Table 4, Theme 5a, explicitly recognize that training was insufficient for many new staff members but, conversely, they also imply that experienced staff (present during the early rollout) were generally satisfied with their training, which explains why this survey finding was determined to be partially consistent with Theme 5a.

Table 4. Qualitative themes on workshop implementation: overarching theme and themes on strengths and weaknesses.

Theme and subtheme	Exemplar quotes	PARIHS[a] construct	Linkages to Table 3 rows
Overarching theme			
1. Importance of outreach and marketing to prompt workshop enrollment	- Local staff, low-enrollment center: "If it was more advertised and marketed, I think that would be good." - Regional leader: "I feel like it's an easy sell. We have the promotional materials to send out." - National leader: "I've learned that it's important to have marketing materials for caregivers. That's a big lesson."	Spans all PARIHS constructs and implementation strengths and weaknesses	N/A[b]
Implementation strengths			
2. Belief in positive impact of workshop encouraged uptake			
2a. Positive caregiver experiences during rollout	- Caregiver: "You learned a lot of really good tools that I use in my daily life. And you could also communicate with the other class members." - Local staff, high-enrollment center: "When the caregivers took the course, I got such positive feedback from them, it made me a believer." - Regional leader: "Anybody that has participated in it has really given lots of positive feedback. And probably the most telling thing is the fact that staff continue to make the referrals."	Evidence—caregiver experiences are positive	Row 1
2b. Value of evidence from prior research[c]	- Local staff, low-enrollment center: "You want something that's, you know, evidence-based." - Regional leader: "The research definitely does matter. Because like I said, you're pitching an additional task to [caregivers] who are super busy."	Evidence—research evidence is convincing	Row 2
3. Successful outreach to some caregiver groups			
3a. Use of stories and testimonials from trusted sources	- Caregiver: "[Caregivers will enroll] if they have good caregiver coordinators that push the program and say, "This is something that you really need to do." Or the therapists, whoever they trust." - Local staff, high-enrollment center: "My suggestion [for other centers] would be short little testimonials of people who've taken the course and it would say, 'See what Mary said about the course.'" - Regional leader: "I always tell people when promoting it—caregivers when they link up they've really enjoyed it."	Evidence—materials help caregiver determine that program is likely to meet their needs; Facilitation—staff have necessary skills	N/A
3b. Multiple contact episodes and materials	- Caregiver: "First I heard verbally about it. Then they sent me a flier regarding it and said, 'Here is what it is – read it and see if this is what you're looking for.' Then he followed up with a phone call. I think the more information, the better." - Local staff, low-enrollment center: "I would say the biggest thing [hindering enrollment] is that we don't have a follow-up plan—a reminder. Because sometimes that reminder helps. To talk about it and you know, give them a little push." - Regional leader: "We've sent out like a large volume of fliers—and I've pitched it to people in person and over the phone and things like that. You can't just do it once. You have to repeatedly send stuff out or bring it up."	Context—staff have sufficient resources; Facilitation—staff have necessary skills	Row 3
Implementation weaknesses and the needs they suggest			
4. Missed opportunities for improved outreach			

Theme and subtheme	Exemplar quotes	PARIHS[a] construct	Linkages to Table 3 rows
4a. Detailed information on workshop content and structure	• Caregiver—who knew about the workshop but had not enrolled: “Provide more actual information on the content, not just a link on the computer.” • Local staff, low-enrollment center: “I don't see how the system works. I would like that. One caregiver was telling me that she didn't feel there was enough information, but without seeing it, I couldn't respond to her.” • Regional leader: “Some of the CSCs[d] do not understand the details. I think it would be helpful to train the coordinators on what actually is in [it]. Because—marketing wise—you're not going to refer people as readily to something that you don't have knowledge about.”	Evidence—caregiver cannot determine if program is likely to meet their needs; Evidence—staff cannot observe program	N/A
4b. Expanded online mechanisms for outreach and enrollment	• Local staff, low-enrollment center: “My suggestion would be a link on the site [portal to electronic health records]—on the same page, if caregivers could just click on right there.” • Local staff, high-enrollment center: “If the general caregivers [caregivers of older, pre-9/11 era veterans] could access it from the caregiver website, that might be really good.” • National leader: “There's a barrier in that you have to pick up the phone and find your CSC. So the recruitment process is interrupted by the fact that it's not a complete online experience, even though the rest of their experience will be online.”	Context—lack of technology tools	N/A
4c. Partnership with communities and community groups	• Only 1 caregiver learned about the workshop through a community group: “Crossroads [VA-funded nonprofit] is where he goes once a month. They just said, ‘Well, try this and maybe they can help you.’ So I picked up a brochure.”[e] • Only 1 local staff member, at a high-enrollment center, described doing outreach to the community: “I do community outreach events where I put out that fact sheet.”[e] • National leader: “We have tremendous access to community-based organizations that are working with caregivers all over the country. And we've wanted to promote the program through that [but have not been able to].”	Context—limited awareness among external networks and communities	N/A
4d. Increased outreach to certain caregiver groups and their health care teams	• Pre-9/11 era caregiver: “I don’t think it’s widely known that you have this. None of the doctors told me about it.” • Local staff, low-enrollment center: “I definitely think outreach and mailers could be made more regularly to the general [pre-9/11 era] caregivers. It just kind of goes by the wayside.” • National leader: “One of the goals is to increase the number of general [pre-9/11 era] caregivers in the program. I don't think the coordinators have as much contact with them.”	Context—limited awareness among some stakeholder groups; Context—limited staff resources and time	Row 3
5. Missed opportunities to support staff in outreach efforts			

Theme and subtheme	Exemplar quotes	PARIHS[a] construct	Linkages to Table 3 rows
5a. Training and mentoring for new staff	• Local staff hired after the rollout began, low-enrollment center: “I had no one to collaborate with, no one to talk to. I think new CSCs need to be more knowledgeable of the program.” • Regional leader: “The initial rollout was either really early in my assuming responsibility for this role or shortly thereafter. Initially it flew right past me.”	Facilitation—lack of training and mentoring from experts	Rows 4 and 5
5b. Improved data management capabilities to • Generate outreach contacts • Track caregivers • Target follow-up outreach	• Local staff, low-enrollment center: “A quarterly flier might just remind people of the opportunity to participate in this program. At the national level can they generate mailers?” • Local staff, high-enrollment center: “It would be very beneficial to have a way to pull up the list of people that we have referred that have never taken any steps forward, so that we can hit a button and they get a reminder email.” • Regional leader: “The website where you can see where the caregiver is in the process, I don't always check it. It probably would be helpful. But you don't always have the time to check individually.” • Regional leader: “Once we make referrals, we’re not involved. Unless there was some sort of process where if the caregiver didn’t follow through, [workshop organizers] came back to the coordinator and said, ‘Will you see if they're still interested?’ There’s no mechanism for that.”	Context—lack of information technology capabilities	Row 6

[a] PARIHS: Promoting Action on Research Implementation in Health Services.

[b] N/A: not applicable.

[c] Caregiver participants did not make comments about research evidence.

[d] CSC: caregiver support coordinator.

[e] These are not exemplar quotes. They are the only cases in which respondents described using these outreach mechanisms and thus highlight their relative absence from use.

Interview Results

During interviews, participants from low-enrollment centers identified fewer implementation strengths and a greater number of weaknesses than those from high-enrollment centers. However, both groups generated similar implementation themes. As the analyses identified common themes, we report findings from low- and high-enrollment centers together.

Theme 1: Importance of Outreach and Marketing to Prompt Workshop Enrollment

We identified an overarching, general theme on the need for effective outreach and marketing to promote workshop enrollment and thereby implementation success (Table 4, theme 1). This theme emerged from comments on implementation strengths and weaknesses and spanned all PARIHS constructs. All participant groups (caregivers, staff, and leaders) emphasized this need:

> *They need to spread the information. Blast it out. Just get the word out.* [Caregiver]

In addition, staff and leaders used marketing language (eg, “market,” “advertise,” “sell,” “promote”) to describe characteristics of the rollout, even in cases where they were not reaching as many caregivers as they would have liked. Of note, none of our survey questions were designed to elicit input on outreach and marketing efforts other than the questions for staff on *informational materials*, as described above.

Respondent feedback also generated themes on specific implementation strengths and weaknesses. Identified themes map onto the PARIHS constructs of evidence, context, and facilitation. They include 2 themes on implementation strengths and 2 on implementation weaknesses, each with respective subthemes.

Theme 2: Belief in Positive Impact of Workshop Encouraged Uptake

All participant groups described an overall positive impact of the workshop. In doing so, they commented on the experiences, needs, and preferences of caregivers as well as research evidence. Both forms of *evidence* were highlighted as important to successful implementation because they made the workshop attractive (to caregivers and staff) and easier to promote.

Subtheme: Positive Caregiver Experiences During Rollout

Caregivers, staff, and leaders identified the workshop as a positive experience for most caregivers (Table 4, subtheme 2a). Caregivers reported that it met expectations of improving their skills and social support, and staff reported receiving similar feedback directly from their local caregivers:

> *In the hallways, they'll stop me saying, “I've been using that online support group, it's really good.”* [Local staff, low-enrollment center]

All groups valued such feedback as evidence of workshop effectiveness.

Subtheme: Evidence From Prior Research

Local staff and regional and national leaders expressed that they also valued research evidence as a motivating factor for their implementation efforts (Table 4, subtheme 2b):

> *Research support matters because it's a waste of time if it's not proven to be effective. And the last thing we want to do is waste caregivers' very precious time.* [Local staff, high-enrollment center]

Research evidence assured them that the workshop was worthwhile for busy caregivers. Caregivers did not mention research evidence in their comments.

Theme 3: Successful Outreach Consisted of Trusted Stories and Multiple Contacts

Another noted strength of the rollout was the achievement of effective outreach to some caregiver groups—specifically, caregivers of younger veterans from the post-9/11 era. These caregivers were best known to local caregiver support coordinators because an act of Congress in 2010 enabled many of them to receive a caregiving stipend from the VA, but payment required quarterly contact with coordinators. In contrast, caregivers of older, pre-9/11 era veterans were not eligible for the stipend, making them less likely to come to the attention of coordinators. Among the post-9/11 caregivers for whom outreach efforts were more successful, certain strategies were identified as important to that success.

Subtheme: Use of Stories and Testimonials From Trusted Sources

Comments from caregivers, local staff, and leaders highlighted that outreach was more effective if it involved caregiver stories and testimonials from trusted sources (Table 4, subtheme 3a):

> *The person who I talked to gave me all the caregiver feedback about its friendly atmosphere. It was a good fit.* [Caregiver]

This type of information was characterized as trustworthy and relatable for caregivers. Knowing that the workshop was helpful to other caregivers prompted them to try it for themselves.

Subtheme: Use of Multiple Contact Episodes and Materials

Caregivers, staff, and leaders described successful outreach as requiring multiple outreach contacts (*touches*) and a variety of formats, for example, in person, fliers, emails, and phone calls (Table 4, subtheme 3b):

> *They teach us in sales that sometimes it takes quite a few touches to convince somebody to even take a look at what you're trying to sell them. We're trying to sell them on the idea of doing this course.* [Local staff, high-enrollment center]

These repeated contacts reminded caregivers that they could enroll in the workshop when they were at a place in their busy lives to be ready and able to do so.

Theme 4: Missed Opportunities for Improved Outreach Limited Workshop Uptake

Caregivers, local staff, and leaders also identified a number of implementation weaknesses. They described missed opportunities for providing effective outreach to caregivers overall and to particular caregiver subgroups, such as caregivers of older, pre-9/11 era veterans, and suggested the need for implementation changes described in the subthemes below.

Subtheme: Detailed Information on Workshop Content and Structure

Some caregivers felt that a lack of detailed information about the workshop on outreach materials inhibited their interest in the program (Table 4, subtheme 4a). Local staff recognized a parallel issue—that their lack of detailed information on the workshop or opportunity to observe it for themselves limited their understanding of it, which constrained their ability to market it:

> *It would be great if we could actually see the information that the caregivers are receiving. Because we could give more feedback on it to caregivers that way—on the modules and what it all entails.* [Local staff, high-enrollment center]

Regional leaders had similar insights on the need to improve staff familiarity with the workshop.

Subtheme: Expanded Online Mechanisms for Outreach and Enrollment

Given that the workshop was web-based, many caregivers were confused by the minimal amount of online information about it and the absence of an online enrollment mechanism. They found these to be barriers to enrollment:

> *When I go on the VA website, there's not too much on the web about it. You got to call this number to find out who the coordinator is, then get in contact. I mean it's so much work to get there.* [Caregiver]

Local staff also desired online options for caregivers to receive information and enroll, particularly caregivers with whom they had little ongoing contact (eg, pre-9/11 era caregivers; Table 4, subtheme 4b). Some leaders agreed that the lack of online mechanisms for outreach and enrollment was an implementation weakness.

Subtheme: Partnership With Communities and Community Groups

Caregivers reported participating in or seeking support from a variety of community groups and community networking sites, many of which were well-established at local, regional, and national levels, such as Veterans of Foreign Wars, Hearts of Valor, the Wounded Warriors Project, Operation Homefront, and Facebook groups for veteran wives. However, there is little evidence to suggest that these community groups were engaged as partners in workshop outreach efforts. Only 1 caregiver reported receiving information on the workshop from a community organization and, in this case, it was a VA-funded nonprofit (Table 4, subtheme 4c). Other caregivers reported hearing about the workshop by word of mouth from veteran spouses who had completed the workshop. Caregivers identified the lack of effective outreach in the community as a missed opportunity:

> *I didn't even know it existed. I had to find it out through somebody I work with. People should know*

> *that this program exists and not just stumble upon it.* [Caregiver]

Similarly, among local staff and regional leaders, only 1 person described performing outreach to community groups. Some leaders recognized community outreach as an important component missing from existing outreach efforts.

Subtheme: Increased Outreach to Certain Caregiver Groups and Their Health Care Teams

Caregivers of older, pre-9/11 veterans expressed insight into the limited outreach efforts made to reach them (Table 4, subtheme 4d). Local staff and leaders made parallel comments about the need to improve targeting of these caregivers as well as other caregiver subgroups (eg, male caregivers of female veterans). They suggested working with health care teams and other support services that commonly interact with older veterans and their caregivers—for example, those in primary care, home health, mental health, respite care, physical and occupational therapy, transportation, and social work outside of the CSP:

> *Partner with the home-based primary care, community-based care, and the PACT [patient-centered medical home] teams. Because that's where you are going to get your older population.* [Regional leader]

Theme 5: Missed Opportunities to Support Staff in Outreach Efforts

Many staff and some regional leaders reported that they did not have the support necessary to perform or maintain effective workshop outreach and indicated the need for improvements described in the subthemes below.

Subtheme: Training and Mentoring for New Staff

Staff who attended the original webinar training on the workshop and rollout gave overall positive feedback (although many still desired to observe the workshop itself, as described above). However, some new staff who began as coordinators after the initial training reported a lack of knowledge, skills, and mentoring. This undermined their motivation and ability to promote the workshop (Table 4, subtheme 5a). Even experienced staff observed this as a problem:

> *I think another national training needs to happen, so that the newer coordinators will see it. Because then I think that they would utilize it.* [Local staff, high-enrollment center]

Leaders also recognized the need for new staff to receive timely and thorough training and mentoring.

Subtheme: Improved Data Management Capabilities

Staff and regional leaders reported having inadequate data management capabilities for many outreach-related activities (Table 4, subtheme 5b). They described ad hoc, inefficient, or nonexistent processes:

> *I think we did have an internal spreadsheet going at one time, but I don't think we've followed up with that.* [Local staff, low-enrollment center]

Staff desired efficient approaches to data processing, especially if they could be performed at the level of the national VA *to generate regular, multiple outreach contacts* to caregivers at their local centers. Second, staff and regional leaders noted that they lacked standardized tools *to track caregivers* along the outreach-to-enrollment pathway. Some staff improvised their own local tracking tools but had trouble maintaining them. Third, the national workshop enrollment database was not integrated with local databases, making it time-consuming to access and infrequently used. These poor data management capabilities undermined the ability of staff *to target follow-up outreach* to specific caregivers. Although local staff felt these deficiencies most acutely, regional leaders also recognized them.

Divergent and Lone Comments

We identified divergent comments from caregivers and staff. Two caregivers expressed that their workshop experience would have been improved by having the other caregivers in their workshop be more like them, according to certain characteristics. One desired to be in a group with younger caregivers, whereas the other wanted to be in a group with caregivers whose care partners had the same medical condition as her care partner. Both recognized beneficial aspects of the existing workshop. A local staff member was supportive of the web-based workshop as a valuable option for caregivers but felt that in-person programs were more beneficial if caregivers could attend them. Finally, another staff member desired to have a Spanish workshop for Spanish-speaking caregivers.

Integrated Findings on Implementation Strengths and Weaknesses

Our study team, advisory board, and a representative from VA CSP leadership determined that the survey and interview findings were consistent with a number of implementation strengths—specifically, that caregiver experiences with the workshop were positive, research evidence supported its use, and staff received some useful outreach materials and resources that helped them support the rollout (Table 3, last column, and Table 4, last column). Findings also converged on implementation weaknesses—that outreach materials and approaches were insufficient in multiple areas, certain staff and regional leaders did not receive necessary training or mentoring, and that data management capabilities were inconsistent or lacking. When we noted areas of apparent inconsistency between quantitative and qualitative results (Table 3, rows 3 and 4), we discussed and noted potential explanations (Table 3, footnotes f and g).

Recommended Strategies for Improving Workshop Uptake

On the basis of the integrated findings, we identified multiple strategies likely to improve workshop uptake and recommended them to VA leadership in the form of a report. The strategies included investment in robust outreach and marketing capabilities, tailoring outreach strategies to all key stakeholder groups and subgroups, use of stories, testimonials, repetition, details, and diverse delivery options in marketing campaigns, recurrent training sessions and mentoring for new staff, and comprehensive data management and reporting capabilities (Table 5).

Table 5. Strategies identified as likely to improve workshop uptake and whether US Department of Veterans Affairs adopted them.

Strategies	Further details	Whether VA[a] adopted
Invest in outreach and marketing capabilities	- May need to contract out	Yes
Tailor specific outreach strategies to all key stakeholder groups	Groups include the following: - Potential workshop participants—may require identifying and targeting important subgroups of caregivers - Local implementation staff with primary responsibility for implementing workshop (ie, caregiver support coordinators) - Other staff who interact with caregivers or their care partners - Community groups	Partially
Use campaigns that are personal, repeated, include workshop details, and have diverse delivery options	- Use personal stories and testimonials—from peers (ie, caregiver peers and staff peers) - Repeat outreach contacts over time—with all key groups, not just caregivers - Include adequate information on workshop details - Enable diverse delivery mechanisms—both electronic and nonelectronic	Yes
Conduct recurrent training and mentoring for new staff	- Is especially important for new staff at locations with no other experienced staff on-site	Yes
Perform comprehensive, ongoing data management and reporting	- To support and track outreach efforts, enrollments, and follow-up on unsuccessful enrollments	Partially

[a] VA: US Department of Veterans Affairs.

Stage 2: Findings on New and Enhanced Implementation Strategies

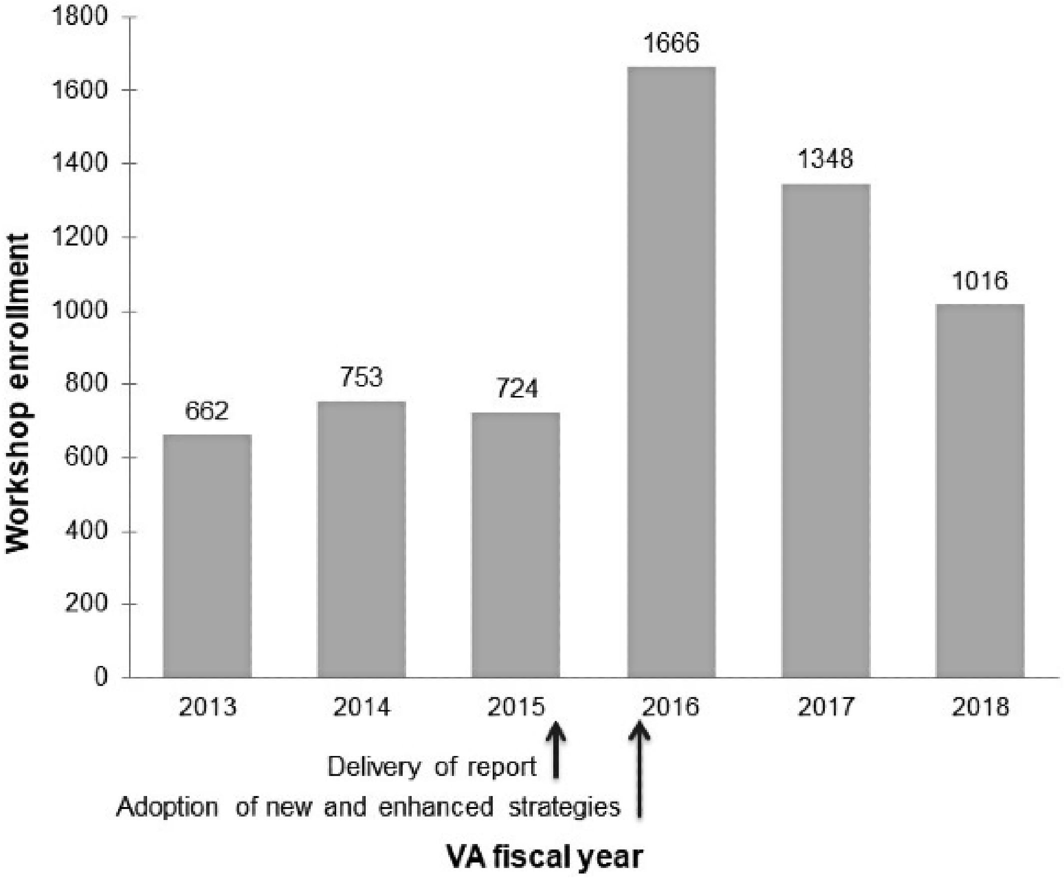
During follow-up interviews in 2018, VA leaders identified multiple implementation changes made in response to stage 1 findings and recommendations (Table 5, last column; see Multimedia Appendix 5 for additional details). The VA CSP contracted with a new organization with greater marketing expertise to administer the workshop, and together they developed and adopted both new and enhanced outreach approaches. They tailored new marketing materials and outreach to implementation staff (eg, new monthly newsletters for caregiver support coordinators), some VA nonimplementation staff and services in close contact with pre-9/11 veterans and their caregivers (eg, primary care social work, home health nursing), and some veteran service organizations and communities (eg, through presentations and press releases). They highlighted stories for caregivers in multiple new ways (eg, videos, quotes, photographs) and did the same for implementation staff (eg, caregiver support coordinators sharing their *BBC stories* with each other in videos or listservs). Outreach campaigns incorporated more repeat *touches* with both groups (eg, regular email invitations to caregivers), diverse forms of delivery (eg, social media posts and phone-based blurbs) and more detailed information on workshop structure and content. For instance, leadership approved and built a demonstration workshop so that staff could learn about workshop details firsthand and then pass the information to the caregivers with whom they worked. Expanded materials for new staff include recurring national training and reminders about how to access outreach resources. Finally, the workshop platform began regularly to email staff with reports on the enrollment status of caregivers at their respective medical centers.

Leaders also reported that they were unable to address some implementation weaknesses. The CSP website, which was managed by a different VA entity with limited funding, was not updated to include more information on the workshop and links to it. The VA national database of registered caregivers remained unlinked from the workshop database, and so it could not be used to generate workshop outreach invitations or other communications to caregivers. Finally, there was no plan to develop a workshop in Spanish.

Longitudinal Workshop Enrollment

Visual inspection of longitudinal enrollment data indicates a possible impact of the new and enhanced implementation strategies. During the initial BBC rollout period (2013 to 2014) and the implementation evaluation period (2015), annual enrollment was approximately 700 caregivers per VA FY (Figure 1). Following implementation changes made in late FY 2015 and early FY 2016, annual workshop enrollment increased. In FYs 2013 to 2015, cumulative enrollment was 2139, which rose to 4030 in FYs 2016 to 2018. Over 6 years, cumulative BBC enrollment reached 6169, which equates to 12% of approximately 50,000 caregivers registered with the VA.

Figure 1. Workshop enrollment, 2013-2018. US Department of VA fiscal years encompass October through the end of September the following year. The study team reported integrated study findings and recommended new and enhanced strategies likely to improve workshop uptake to VA Building Better Caregivers national leaders in the late fiscal year 2015. In late 2015 and early 2016, the VA adopted a number of these recommended implementation strategies. VA: Veterans Affairs.



Discussion

Principal Findings

This study expands the limited implementation science literature on best practices for health systems to use when implementing web-based psychoeducational programs [13,31]. In our evaluation of the BBC workshop rollout among VA centers nationwide, we identified implementation strengths and weaknesses, including belief in the positive impact of the workshop, use of some successful outreach approaches for caregivers, and missed opportunities to improve outreach and to support both new and experienced implementation staff. On the basis of these findings, we identified and recommended to VA leadership new or enhanced implementation strategies likely to improve workshop uptake—specifically, increased investment in outreach and marketing capabilities, tailoring outreach strategies to key stakeholder groups and subgroups, use of campaigns that are personal, repeated, detailed, and have diverse delivery options, recurrent training and mentoring for new staff, and comprehensive data management and reporting capabilities to support staff implementation efforts. We determined that the VA adopted many of these strategies and, in the years following adoption, national BBC enrollment increased.

Limitations

Our evaluation has a number of limitations. We initiated this after the workshop rollout was underway, rather than prospectively. At the same time, this enabled us to quickly identify low- and high-enrollment sites for inclusion in the evaluation. Our evaluation is also limited to a single health care system. Nonetheless, the VA health care system is the largest integrated health care system in the United States and, as such, it has numerous medical centers. These demonstrated variation in implementation success as well as geography and catchment size, which enhanced the robustness of our evaluation. The VA workshop is available only in English, which precludes use by caregivers who have limited English proficiency.

Comparison With Prior Work

Among caregiver programs, many have been funded as research projects but few, if any, have been the focus of published implementation evaluations. A recent systematic review identified 17 studies of web-based caregiver programs focused on education, health, and well-being [7]. However, we were unable to identify any with a published implementation evaluation. Among web-based caregiver programs identified by other means [32, 33], no implementation evaluations were found. Finally, we identified a process evaluation of a web-based intervention to support frail older adults that, at their discretion, could also involve their caregivers but was not designed for caregivers [34]. Thus, it remains to be seen how results from the current implementation evaluation may compare with findings from implementation evaluations of other web-based psychoeducational programs for caregivers.

Among caregivers who care for military veterans, US VA caregivers may be unique in the amount of support they receive. Since 2010, the VA and VA CSP have received substantial federal funding for caregiver programs and research. This has enabled the US VA and VA investigators to develop the BBC workshop and other VA caregiver programs (eg, Resources for Enhancing All Caregivers Health [REACH] VA); assess such programs for efficacy; fund their implementation; and in the case of this study, evaluate their implementation [22,35]. We are unaware of funding opportunities on the same scale in other countries. This may explain why evidence on the efficacy and implementation of veteran caregiver programs outside the United States is limited. In an environmental scan of resources (defined as programs and services) for families of veterans with operational stress injuries, investigators identified 42 non–United States resources but concluded, "Information on evaluation and evidence for resources is limited" [36]. To our knowledge, none were the focus of an implementation evaluation.

If one examines web-based or technology-based interventions for patients rather than caregivers, implementation data remain sparse [4,31,34,37,38]. A 2013 systematic review of studies of web-based patient self-management workshops determined that 87% "did not describe any measures taken to sustain the tested programs past designated study time periods" [31]. The remainder reported on the number of participants reached but not the strategies that were successful or unsuccessful in achieving participation. Not surprisingly, a similar 2013 review of web-based interventions for psychosocial health determined that issues of reach and adoption of web-based interventions were a crucial area for further research [4]. Four years later, a 2017 review of 268 randomized controlled trials described a proliferation of evidence-based self-guided internet health programs but concluded, "Unfortunately, most of the Internet delivered health interventions that were efficacious through RCTs were not available after the conclusion of the trials" [3] and we could not identify implementation evaluations of these same trials. Finally, in the case of eHealth services broadly defined (eg, telehealth, patient portals, web-based decision aids), others have identified a taxonomy of implementation barriers and facilitators from multiple studies and expert input [11], but there are few follow-up studies to determine whether making implementation enhancements suggested by the taxonomy does indeed boost uptake. Thus, experts have highlighted implementation research as the crucial next step in spreading the benefits of evidence-based digital health interventions to all those who might benefit [39].

This evaluation suggests that implementation research can help boost uptake of web-based psychoeducational interventions for caregivers and perhaps others. Among the results, we were most surprised by the prominence of findings related to outreach and marketing. The implementation science literature has not emphasized these concepts to any great extent, although this may be changing. This evaluation was performed before the publication of the most recent iteration of the PARIHS implementation framework known as the i-PARIHS framework [40]. i-PARIHS adds a new construct, "recipients" to existing constructs (evidence, context, and facilitation) and defines it as "the people who are affected by and influence implementation [of the intervention] at both the individual and collective team level" [40]. In this study, we found implementation could be strengthened when specific groups and subgroups of *recipients* (eg, caregivers, staff, community organizations) were explicitly identified for tailored outreach campaigns, which aligns with this new PARIHS construct. Implementation scientists have started to note that marketplace principles and storytelling may contribute to implementation success [41,42]. Similarly, the recent report on a US National Cancer Institute implementation accelerator program concluded that "researchers can benefit from learning the 'language' of business" when "trying to move their research from the lab to the real world" [43].

Conclusions

In this mixed methods implementation evaluation of the web-based psychoeducational BBC workshop for US veteran caregivers, we identified initial implementation strengths and weaknesses as well as multiple strategies likely to improve workshop uptake—including strong investment in outreach and marketing capabilities, tailoring of specific outreach strategies to all key stakeholder groups, use of outreach campaigns that are personal, repeated, detailed, and have diverse delivery options, ongoing training and mentoring for new implementation staff, and comprehensive data management and reporting capabilities. Upon follow-up assessment, we determined that the VA enhanced or newly adopted a number of these recommended strategies nationally and that workshop enrollment subsequently increased. Other health systems may want to deploy these strategies when implementing their web-based programs.

Acknowledgments

VY conceived and designed the study, designed data collection instruments, interviewed participants, performed qualitative analyses as part of the qualitative evaluation team, interpreted findings, wrote the manuscript, and guaranteed the integrity of the work as a whole. RG (deceased) helped to design the interview guides and modify the survey, interviewed participants, performed qualitative analyses as part of the qualitative evaluation team, contributed to interpretation of findings, managed the study, and revised an earlier draft of the manuscript critically for important intellectual content. AN helped design the interview guides, interviewed participants, performed qualitative analyses as part of the qualitative evaluation team, contributed to interpretation of findings, and revised the manuscript critically for important intellectual content. LO helped design the interview guides, performed caregiver participant recruitment, managed study data, and revised the manuscript critically for important intellectual content. CK contributed to the design of the study and interview guides, performed qualitative analyses as part of the qualitative evaluation team, and revised the manuscript critically for important intellectual content. RT contributed to the design of the study and interpretation of findings and revised the manuscript critically for important intellectual content. ND contributed to the acquisition of data and interpretation of findings and revised the manuscript critically for important intellectual content. MK contributed to the acquisition of data and interpretation of findings and revised the manuscript critically for important intellectual content. SA contributed to the design of the study and the interpretation of findings and revised the manuscript critically for important intellectual content.

This study was supported by a grant from the VA Quality Enhancement Research Initiative (Project 14-195-1). The funder was not involved in the design of this study, the collection, analysis, and interpretation of the data or the writing of the manuscript. The views expressed herein are those of the authors and do not necessarily reflect the position or policy of the US Department of Veterans Affairs.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guides used for phone interviews of caregivers, staff, and leaders.

[DOC File , 82 KB - jmir_v23i1e16495_app1.doc]

Multimedia Appendix 2

Caregiver paper-based survey.

[PDF File (Adobe PDF File), 181 KB - jmir_v23i1e16495_app2.pdf]

Multimedia Appendix 3

Staff online survey.

[PDF File (Adobe PDF File), 138 KB - jmir_v23i1e16495_app3.pdf]

Multimedia Appendix 4

Available US Department of Veterans Affairs operations data on characteristics of caregivers enrolled in BBC workshop during rollout years 1-2 and their veteran care partners.

[DOCX File , 16 KB - jmir_v23i1e16495_app4.docx]

Multimedia Appendix 5

Detailed description of implementation changes made by US Department of Veterans Affairs in response to stage 1 findings.

[DOCX File , 18 KB - jmir_v23i1e16495_app5.docx]

References

1. Lorig K, Ritter PL, Plant K, Laurent DD, Kelly P, Rowe S. The South Australia health chronic disease self-management internet trial. Health Educ Behav 2013 Feb;40(1):67-77. [doi: 10.1177/1090198112436969] [Medline: 22491008]
2. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
3. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
4. Paul CL, Carey ML, Sanson-Fisher RW, Houlcroft LE, Turon HE. The impact of web-based approaches on psychosocial health in chronic physical and mental health conditions. Health Educ Res 2013 Jun;28(3):450-471 [FREE Full text] [doi: 10.1093/her/cyt053] [Medline: 23660463]
5. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
6. Zhou T, Li X, Pei Y, Gao J, Kong J. Internet-based cognitive behavioural therapy for subthreshold depression: a systematic review and meta-analysis. BMC Psychiatry 2016 Oct 21;16(1):356 [FREE Full text] [doi: 10.1186/s12888-016-1061-9] [Medline: 27769266]
7. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]

8. Rost T, Stein J, Löbner M, Kersting A, Luck-Sikorski C, Riedel-Heller SG. User acceptance of computerized cognitive behavioral therapy for depression: systematic review. J Med Internet Res 2017 Sep 13;19(9):e309 [FREE Full text] [doi: 10.2196/jmir.7662] [Medline: 28903893]
9. Coffey NT, Cassese J, Cai X, Garfinkel S, Patel D, Jones R, et al. Identifying and understanding the health information experiences and preferences of caregivers of individuals with either traumatic brain injury, spinal cord injury, or burn injury: a qualitative investigation. J Med Internet Res 2017 May 10;19(5):e159 [FREE Full text] [doi: 10.2196/jmir.7027] [Medline: 28490418]
10. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
11. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
12. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
13. Gitlin L, Marx K, Stanley I, Hodgson N. Translating evidence-based dementia caregiving interventions into practice: state-of-the-science and next steps. Gerontologist 2015 Apr;55(2):210-226 [FREE Full text] [doi: 10.1093/geront/gnu123] [Medline: 26035597]
14. Teno JM, Price RA, Makaroun LK. Challenges of measuring quality of community-based programs for seriously ill individuals and their families. Health Aff (Millwood) 2017 Jul 1;36(7):1227-1233. [doi: 10.1377/hlthaff.2017.0161] [Medline: 28679809]
15. Schulz R, Beach SR, Friedman EM, Martsolf GR, Rodakowski J, James AE. Changing structures and processes to support family caregivers of seriously ill patients. J Palliat Med 2018 Mar;21(S2):S36-S42 [FREE Full text] [doi: 10.1089/jpm.2017.0437] [Medline: 29091533]
16. Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, et al. Caregiver-focused, web-based interventions: systematic review and meta-analysis (part 2). J Med Internet Res 2018 Oct 26;20(10):e11247 [FREE Full text] [doi: 10.2196/11247] [Medline: 30368439]
17. Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of internet-based interventions on caregiver mental health: systematic review and meta-analysis. J Med Internet Res 2018 Jul 3;20(7):e10668 [FREE Full text] [doi: 10.2196/10668] [Medline: 29970358]
18. Barbabella F, Poli A, Andréasson F, Salzmann B, Papa R, Hanson E, et al. A web-based psychosocial intervention for family caregivers of older people: results from a mixed-methods study in three european countries. JMIR Res Protoc 2016 Oct 6;5(4):e196 [FREE Full text] [doi: 10.2196/resprot.5847] [Medline: 27713113]
19. Cheng S, Li K, Losada A, Zhang F, Au A, Thompson LW, et al. The effectiveness of nonpharmacological interventions for informal dementia caregivers: an updated systematic review and meta-analysis. Psychol Aging 2020 Feb;35(1):55-77. [doi: 10.1037/pag0000401] [Medline: 31985249]
20. Caregivers of Veterans – Serving on the Homefront. United Health Group. 2010 Nov. URL: https://www.unitedhealthgroup.com/content/dam/UHG/PDF/uhf/caregivers-of-veterans-study.pdf [accessed 2020-12-09]
21. Ramchand R, Tanielian T, Fisher M, Vaughan C, Trail T, Epley C, et al. Hidden Heroes: America's Military Caregivers. Santa Monica, CA: RAND Corporation; 2014.
22. Dupke NJ, Plant KL, Kosteas J. Supporting caregivers of veterans online: a partnership of the national council on aging and VA. Fed Pract 2016 Jan;33(1):41-46 [FREE Full text] [Medline: 30766137]
23. Lorig K, Thompson-Gallagher D, Traylor L, Ritter PL, Laurent DD, Plant K, et al. Building Better Caregivers: a pilot online support workshop for family caregivers of cognitively impaired adults. J Appl Gerontol 2010 Dec 15;31(3):423-437. [doi: 10.1177/0733464810389806]
24. Lorig K, Ritter PL, Laurent DD, Yank V. Building better caregivers: a pragmatic 12-month trial of a community-based workshop for caregivers of cognitively impaired adults. J Appl Gerontol 2019 Sep;38(9):1228-1252 [FREE Full text] [doi: 10.1177/0733464817741682] [Medline: 29165000]
25. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
26. Rycroft-Malone J. The PARIHS framework-a framework for guiding the implementation of evidence-based practice. J Nurs Care Qual 2004;19(4):297-304. [doi: 10.1097/00001786-200410000-00002] [Medline: 15535533]
27. Stetler CB, Damschroder LJ, Helfrich CD, Hagedorn HJ. A Guide for applying a revised version of the PARIHS framework for implementation. Implement Sci 2011 Aug 30;6:99. [doi: 10.1186/1748-5908-6-99] [Medline: 21878092]
28. Hagedorn HJ, Heideman PW. The relationship between baseline organizational readiness to change assessment subscale scores and implementation of hepatitis prevention services in substance use disorders treatment clinics: a case study. Implement Sci 2010 Jun 14;5:46 [FREE Full text] [doi: 10.1186/1748-5908-5-46] [Medline: 20546584]
29. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]

30. Neal JW, Neal ZP, van Dyke E, Kornbluh M. Expediting the analysis of qualitative data in evaluation. Am J Eval 2014 May 27;36(1):118-132. [doi: 10.1177/1098214014536601]
31. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
32. Blusi M, Dalin R, Jong M. The benefits of e-health support for older family caregivers in rural areas. J Telemed Telecare 2014 Mar;20(2):63-69. [doi: 10.1177/1357633X13519901] [Medline: 24446251]
33. Griffiths PC, Kovaleva M, Higgins M, Langston AH, Hepburn K. Tele-savvy: an online program for dementia caregivers. Am J Alzheimers Dis Other Demen 2018 Aug;33(5):269-276 [FREE Full text] [doi: 10.1177/1533317518755331] [Medline: 29544342]
34. Robben SH, Perry M, Huisjes M, van Nieuwenhuijzen L, Schers HJ, van Weel C, et al. Implementation of an innovative web-based conference table for community-dwelling frail older people, their informal caregivers and professionals: a process evaluation. BMC Health Serv Res 2012 Aug 15;12:251 [FREE Full text] [doi: 10.1186/1472-6963-12-251] [Medline: 22894654]
35. Nichols LO, Martindale-Adams J, Burns R, Graney MJ, Zuber J. Translation of a dementia caregiver support program in a health care system-REACH VA. Arch Intern Med 2011 Feb 28;171(4):353-359. [doi: 10.1001/archinternmed.2010.548] [Medline: 21357811]
36. Tam-Seto L, Cramm H, Norris D, Eichler M, Smith-Evans K. An environmental scan of programs and services for families of veterans with operational stress injuries. Mil Behav Health 2016 May 9;4(4):390-397. [doi: 10.1080/21635781.2016.1187098]
37. Hermes E, Burrone L, Perez E, Martino S, Rowe M. Implementing internet-based self-care programs in primary care: qualitative analysis of determinants of practice for patients and providers. JMIR Ment Health 2018 May 18;5(2):e42 [FREE Full text] [doi: 10.2196/mental.9600] [Medline: 29776898]
38. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Emental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
39. Buis L. Implementation: the next giant hurdle to clinical transformation with digital health. J Med Internet Res 2019 Nov 20;21(11):e16259 [FREE Full text] [doi: 10.2196/16259] [Medline: 31746763]
40. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11:33 [FREE Full text] [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]
41. King DK, Shoup JA, Raebel MA, Anderson CB, Wagner NM, Ritzwoller DP, et al. Planning for implementation success using RE-AIM and CFIR frameworks: a qualitative study. Front Public Health 2020;8:59 [FREE Full text] [doi: 10.3389/fpubh.2020.00059] [Medline: 32195217]
42. Brownson RC, Fielding JE, Green LW. Building capacity for evidence-based public health: reconciling the pulls of practice and the push of research. Annu Rev Public Health 2018 Apr 1;39:27-53 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-014746] [Medline: 29166243]
43. Oh A, Gaysynsky A, Knott C, Nock N, Erwin D, Vinson CA. Customer discovery as a tool for moving behavioral interventions into the marketplace: insights from the NCI SPRINT program. Transl Behav Med 2019 Nov 25;9(6):1139-1150 [FREE Full text] [doi: 10.1093/tbm/ibz103] [Medline: 31313817]

Abbreviations

BBC: Building Better Caregivers
CSP: Caregiver Support Program
FY: fiscal year
ORCA: Organizational Readiness to Change Assessment
PARIHS: Promoting Action on Research Implementation in Health Services
VA: US Department of Veterans Affairs

Edited by G Eysenbach; submitted 03.10.19; peer-reviewed by A McKinlay, S Ge, A Durante; comments to author 23.03.20; revised version received 13.07.20; accepted 18.11.20; published 07.01.21.

Please cite as:
Yank V, Gale RC, Nevedal A, Okwara L, Koenig CJ, Trivedi RB, Dupke NJ, Kabat M, Asch SM
Improving Uptake of a National Web-Based Psychoeducational Workshop for Informal Caregivers of Veterans: Mixed Methods Implementation Evaluation
J Med Internet Res 2021;23(1):e16495
URL: https://www.jmir.org/2021/1/e16495
doi:10.2196/16495
PMID:33410759

Original Paper

Lifelog Data-Based Prediction Model of Digital Health Care App Customer Churn: Retrospective Observational Study

Hongwook Kwon[1], MS; Ho Heon Kim[1], MS; Jaeil An[1], MS; Jae-Ho Lee[2,3], MD, PhD; Yu Rang Park[1], PhD

[1]Department of Biomedical Systems Informatics, College of Medicine, Yonsei University, Seoul, Republic of Korea
[2]Department of Information Medicine, Asan Medical Center, College of Medicine, University of Ulsan, Seoul, Republic of Korea
[3]Department of Emergency Medicine, Asan Medical Center, College of Medicine, University of Ulsan, Seoul, Republic of Korea

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
College of Medicine
Yonsei University
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 2493
Email: yurangpark@yuhs.ac

Abstract

Background: Customer churn is the rate at which customers stop doing business with an entity. In the field of digital health care, user churn prediction is important not only in terms of company revenue but also for improving the health of users. Churn prediction has been previously studied, but most studies applied time-invariant model structures and used structured data. However, additional unstructured data have become available; therefore, it has become essential to process daily time-series log data for churn predictions.

Objective: We aimed to apply a recurrent neural network structure to accept time-series patterns using lifelog data and text message data to predict the churn of digital health care users.

Methods: This study was based on the use data of a digital health care app that provides interactive messages with human coaches regarding food, exercise, and weight logs. Among the users in Korea who enrolled between January 1, 2017 and January 1, 2019, we defined churn users according to the following criteria: users who received a refund before the paid program ended and users who received a refund 7 days after the trial period. We used long short-term memory with a masking layer to receive sequence data with different lengths. We also performed topic modeling to vectorize text messages. To interpret the contributions of each variable to model predictions, we used integrated gradients, which is an attribution method.

Results: A total of 1868 eligible users were included in this study. The final performance of churn prediction was an F1 score of 0.89; that score decreased by 0.12 when the data of the final week were excluded (F1 score 0.77). Additionally, when text data were included, the mean predicted performance increased by approximately 0.085 at every time point. Steps per day had the largest contribution (0.1085). Among the topic variables, poor habits (eg, drinking alcohol, overeating, and late-night eating) showed the largest contribution (0.0875).

Conclusions: The model with a recurrent neural network architecture that used log data and message data demonstrated high performance for churn classification. Additionally, the analysis of the contribution of the variables is expected to help identify signs of user churn in advance and improve the adherence in digital health care.

(J Med Internet Res 2021;23(1):e22184) doi:10.2196/22184

KEYWORDS

churn prediction; digital health care; life-log data; topic modeling; recurrent neural network; deep learning interpretation; attribution method; integrated gradients; digital health; prediction; model; data; app; observational; time-series; neural network

Introduction

Customer churn prediction is one of the most important concerns for almost every company. If customers leave, then sales are reduced, and new customers are needed to replace the previous ones [1]. However, the cost of attracting new customers is 5- to 10-times higher than the cost of retaining customers [2]; therefore, it is much more effective to predict potential churn and prevent these customers from leaving by utilizing promotions or marketing.

Digital health care refers to public health services ranging from simple weight management to professional medicine using mobile devices [3]. With smartphone use becoming more common, the digital health care industry is growing, and numerous health-related apps have been launched [4]. This has provided many people with convenient access to digital health care; however, for them to achieve actual improvements in health, they need to use the apps consistently [5,6]. Therefore, the prediction of churn and the retention of digital health care service customers have significant implications for companies and for users.

Because of the importance of predicting customer churn, studies [7,8] have been performed. However, these studies have generally been conducted using statistical methods and time-insensitive machine learning techniques (eg, decision tree, logistic regression, or support vector machine) [7], by which some information can be lost during the preprocessing sequence [8]. Therefore, a model structure that can accept time-series patterns is necessary.

Many studies have used structured information about customers, which is generally stored in customer relationship management databases [9]. However, more customer text data are becoming available, such as online posts and messages, and it is known that analyzing text data improves predictive performance in customer churn problems [10,11].

We aimed to apply a recurrent neural network structure to leverage time-series patterns in user lifelog data and text message data to predict user churn for digital health care apps. We also aimed to examine the impact of time-series data on model performance and whether the presence of text data affects the performance of churn prediction.

Methods

Health Care App

This study was conducted using data from Noom (Noom Inc), a global digital health care app that provides lifestyle-related functions, such as food logs, exercise logs, weight logs, in-app group activities, and in-app articles. Users are encouraged to log their food intake, exercise every day, and record their weight every week [12]. Users of this digital health care service can also send messages to personal coaches to ask questions about dietary intake, exercise, mindset, or program descriptions. Personal coaches offer feedback to users, in the form of praise, emotional support, encouragement, and validation, based on the user's entries [12].

Study Design

This study predicted customer churn based on lifelog data provided during the customer’s use period. Both paid and free services are available in this digital health care app, and the study defined customers who were refunded for their paid services as churn users.

We received anonymized and unidentified payment information and log data from the company for Korean users only. User data were randomly selected among users who had service payment records between January 1, 2017 and January 1, 2019. Users can select a program lasting 4, 8, 16, or 24 weeks; we limited our analysis to the 16-week program because the largest number of customers chose this period. Additionally, the service use period of the program varied depending on the date of payment and whether the users churned; therefore, we used only log data corresponding to each user’s paid service period for analysis.

Because the proportion of customers who request a refund is generally very small compared with the proportion of total users, the same ratio of retained users was extracted to match the number of churn users to address the problem of data imbalance associated with machine learning [13]. First, users who cancelled their subscriptions, with the refund date recorded, were selected. If their refund occurred after the end of the 16-week program, users were considered retained users and were excluded from the churn user group. The paid program includes a trial period lasting 7 days; during that time, users can request a refund after the initial payment. Therefore, users who received refunds during the trial period were excluded from the churn group because data for fewer than 7 days are insufficient for analysis reliability and generalization. Finally, 934 users were included as the churn group, and 934 retained customers were randomly selected for inclusion as the retained group; the 2 groups were equally matched with respect to gender [14].

This study was approved by the institutional review board (2017-1253) of the Asan Medical Center. The need for informed consent was waived because this study used routinely collected log data that were anonymously managed at all stages, including during data cleaning and statistical analyses.

Model Structure

The overall predictive model structure was designed to include both time-variant and time-invariant inputs. Inputs that occur over time include lifelog data such as step records, daily weight measurement records, diet intake records, and user text messages. Inputs that were considered constant included age, sex, initial BMI, or target BMI of each user (Figure 1).

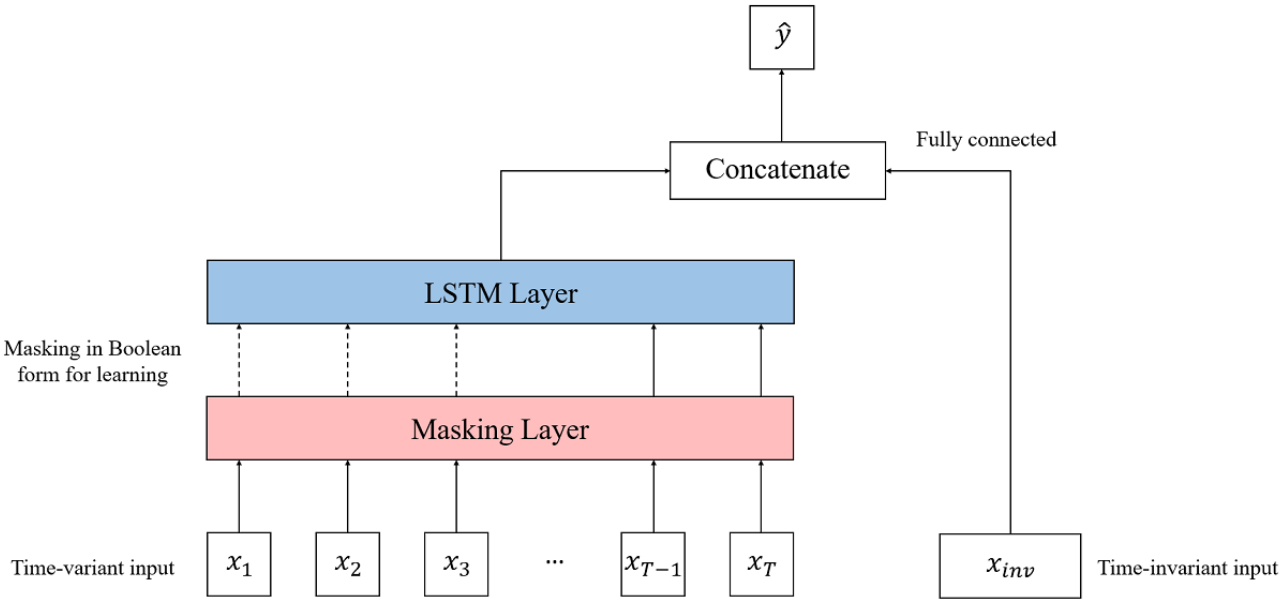
The time-variant node used a long short-term memory structure, which belongs to the recurrent neural network family and is good for processing time-series data. Long short-term memory was developed to solve the vanishing gradient problem that can occur when training a basic recurrent neural network [15,16]. Long short-term memory has a memory cell that contains a node with a self-connected recurrent edge of a fixed weight, thereby ensuring that the gradient can exist over long time steps without vanishing or exploding [17].

Long short-term memory, as in other recurrent neural network-based models, should have the same sequence length input for time; therefore, the time steps of every sample need to be adjusted to ensure that they are the same. The period of service use was constant for retained users; however, for those in the churn group, the period of service varied depending on the time of departure of the user. Therefore, zero padding was used to lengthen the service period data of the churn group to 16 weeks.

One possible critical problem is that the day data of churn users are padded with zeros. Another is that the actual existing data of churn users are relatively short compared with the data of the retained users. If our algorithms notice that part of the sequence is padded with zeros (ie, in the case of churn users), then the problem of leaking the actual labels occurs. A masking layer was added to keep the model unaware of the length of the input sequence. The masking layer produces a Boolean-type tensor for learning whether to use certain values or to ignore them in downstream tasks, and it is used to handle time-series data of different lengths [18].

After padding and masking, time-variant inputs were processed in the long short-term memory layer and the outputs of the layer were concatenated with the time-invariant inputs. Finally, binary classification of the churn was performed.

Figure 1. Overall model architecture. LSTM: long short-term memory.



Text Message Preprocessing

The Noom app provides a function for users and coaches to communicate with each other through text messages. The messages that users receive from coaches can vary depending on their assigned coach; therefore, only messages sent by users were included in this study. User intentions can be more directly inferred from text messages than from log data. To input text message data in the model, the messages were vectorized. Although there are many advanced word-level and sentence-level embedding techniques, topic modeling with latent Dirichlet allocation was used in this study.

Topic modeling assumes each document is a set of random topics and probabilistically presents the importance of the topics and words in the document [19]. Latent Dirichlet allocation was used to estimate the probability that a word corresponds to a particular topic and the probability that a particular topic exists to find the topics in each unstructured document [20]. After topic modeling, each topic name was labeled as intended by the researcher (ie, the name of the topic is not determined by any criteria, but rather by discretion), taking into account the distribution of words corresponding to each topic. Topic modeling has the advantage of easy implementation and intuitive interpretation by examining the proportion of topics in sentences. For each text message, topic modeling was conducted to demonstrate the distribution of the topic proportion of each message. Additionally, since one person could send multiple messages during 1 day, the means of topic vectors was calculated to check the overall topic of the day, and the maximum of the vectors was taken to ensure that important topics that do not appear frequently were not diluted in the average.

Although topic modeling is popular because it is simple to use and easy to understand, there is a limitation: researchers must determine the optimal number of topics present in the documents [21]. This is not a serious problem if the exact number of topics is known, but it is very difficult to select the optimal number of topics without prior knowledge. To solve this problem, the optimal number of topics was determined using the coherence score, which calculates the similarity between the words included in the topic and calculates whether the topic consists of words that have semantic similarity [22].

Interpretation Method

Identifying signs of churn in advance is extremely important for a company; however, because of its nested nonlinear structure, most deep learning models, including our model, are black boxes that, despite their good performance, do not provide information regarding the basis of the predictions [23]. To compensate for this uninterpretability, integrated gradients, which are part of an attribution method, were applied.

Attribution methods produce explanations of an analytic model by assigning an attribution value (relevance or contribution) to each input feature [24]. We used integrated gradients—a final prediction is calculated by multiplication of each variable and coefficient, and the output of the model is also a product of inputs and gradients in the deep learning structure, similar to linear regression [25]. The integrated gradients method was designed to improve the simple gradient approach, which does not satisfy implementation invariance (ie, the attributions are always identical for 2 functionally equivalent networks). Integrated gradients attempt to capture the effects of nonlinearities by computing the gradient along a line between the input data and given reference baseline data [26]. The integrated gradient of the dimension i is defined as follows:

where

which represents a deep neural network function; $x \in R^n$ represents the input; and $x' \in R^n$ represents the baseline data (eg, zero-embedding vector for text neural network) [27]. Using this method, we can examine the effect of each variable on the final output of the model; to explain the effect of each variable on the model, we investigated the average value of the integrated gradients for each variable.

Model Evaluation

During the model performance evaluation, we compared the performance from 2 perspectives: (1) how the performance of the model changes as data close to the point of churn (or retention) is determined at the last date, with some daily data excluded sequentially from the end of the period and (2) the performance differences in the model depending on the data regarding the presence of the text message vectors at every time point. Classification accuracy, F1 score, and area under the receiver operating characteristic curve were calculated.

This study was implemented with Python programming language (version 3.6.8). Data preprocessing procedures related to topic modeling were implemented using the gensim package (version 3.8.3). All neural network modeling including long short-term memory and masking layer were implemented using TensorFlow (version 1.14.0) and Keras (version 2.3.1).

Results

Baseline Characteristics

A total of 1868 eligible users (934 churn and 934 retained), were included in this study. The churn and retained groups showed no statistical differences in gender (*P*>.999) and age (*P*=.20). Both groups comprised mostly women (both groups 825/934, 88.3%), and the mean age was approximately 31 years. The initial BMI and target BMI were calculated based on the height, initial weight, and weight targeted by service use, and both showed no statistically significant differences (*P*=.41 and *P*=.19, respectively). Statistically significant differences were found for the total service period (*P*<.001) and the number of input days for meals (*P*<.001), messages (*P*<.001), walking (*P*<.001), and weight tables (*P*<.001), which are time-variant log data (Table 1).

Table 1. Demographic characteristics.

Variables	Churn users (n=934)	Retained users (n=934)	*P* value
Gender, n (%)			>.999
Female	825 (88.3)	825 (88.3)	
Male	109 (11.7)	109 (11.7)	
Age (years), mean (SD)	31.3 (7.1)	31.7 (7.9)	.20
Initial BMI, kg/m^2, mean (SD)	25.2 (4.1)	25.1 (3.8)	.41
Target BMI, kg/m^2, mean (SD)	22.0 (3.6)	21.8 (3.4)	.19
Total number of days of service use, mean (SD)	43.9 (31.3)	112.0 (N/A[a])	<.001
Meal input days	24.0 (23.2)	53.8 (35.8)	<.001
Message sent days	11.8 (11.1)	22.3 (16.9)	<.001
Walk days	40.3 (29.5)	82.9 (32.9)	<.001
Weigh-in days	9.2 (11.4)	20.6 (19.9)	<.001

[a] N/A: not applicable.

Topic Modeling

The value of coherence according to the number of topics is shown in Figure 2. The scores steadily increased for up to 9 topics (0.6469); then, they repeatedly fluctuated and did not show much further increase. Therefore, the optimal number of topics was determined to be 9.

The results—the weighted proportion of the top 10 words—of topic modeling with 9 topics are shown in Table 2.

Figure 2. Coherence score by the number of topics.

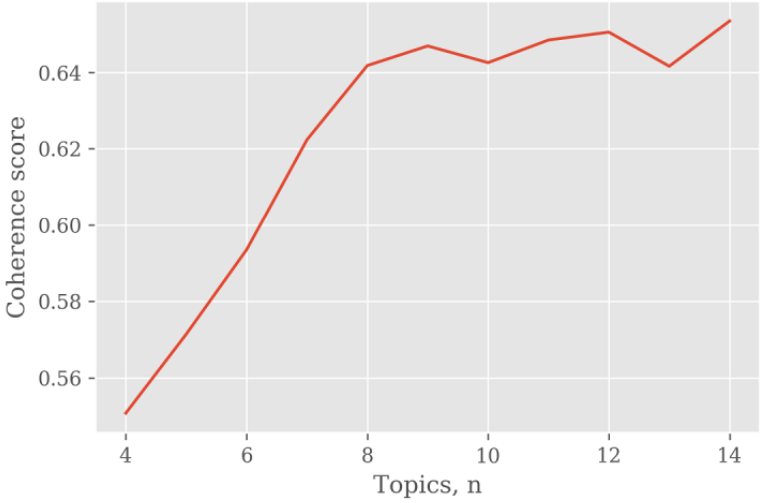


Table 2. Topic modeling results using a combination of the top 10 keywords and weight.

Top 10	Topic								
	Food plan	Schedule	Dining plan	Bad habit	Weight management	Daily life	Dietary intake	General exercise	Coaching
Rank 1									
Keyword	Dinner	Today	Record	Water	Weekend	Time	Thought	Exercise	Thanks
Weight	0.079	0.163	0.088	0.042	0.068	0.085	0.067	0.282	0.114
Rank 2									
Keyword	Today	Yesterday	Tomorrow	Alcohol	Weight	Morning	Effort	Muscle	Diet
Weight	0.032	0.065	0.05	0.036	0.065	0.074	0.052	0.021	0.032
Rank 3									
Keyword	Meal	Next	Hello	Weirdness	Coach	Dinner	Control	Squat	Health
Weight	0.031	0.027	0.028	0.035	0.064	0.059	0.051	0.02	0.026
Rank 4									
Keyword	Snack	Saturday	Day	Weight	Goal	Lunch	Calorie	Stretching	Preparation
Weight	0.03	0.025	0.028	0.026	0.052	0.05	0.049	0.02	0.026
Rank 5									
Keyword	Promise	Start	Travel	Binge	Weight reduction	Meal	Food	Home	Coach
Weight	0.03	0.024	0.023	0.025	0.037	0.04	0.044	0.017	0.023
Rank 6									
Keyword	This time	Mission	Meal	Late-night meal	Height	Leave work	Diet	Aerobic	Stress
Weight	0.028	0.023	0.02	0.017	0.036	0.02	0.043	0.013	0.022
Rank 7									
Keyword	Home	Coach	Menu	Mind	Now	Company	Maximum	Gym	Body
Weight	0.025	0.021	0.018	0.015	0.02	0.018	0.016	0.012	0.021
Rank 8									
Keyword	Plan	Sunday	Eat out	Stamina	This time	Home	Intake	Feeling	Finish
Weight	0.025	0.02	0.017	0.014	0.017	0.016	0.014	0.012	0.021
Rank 9									
Keyword	Weekend	Action	Business dinner	Night	Hello	Possible	Fat	Week	Concern
Weight	0.025	0.02	0.015	0.014	0.016	0.015	0.013	0.012	0.02
Rank 10									
Keyword	Salad	This time	Person	Input	Management	Go to work	Meal	Start	Habit
Weight	0.021	0.02	0.014	0.014	0.016	0.012	0.013	0.011	0.02

Model Performance

The results shown in Table 3 show the classification accuracy, F1 score, and area under the receiver operating characteristic curve for the same test data when (1) some daily data were excluded sequentially from the end of the period and (2) with or without the text message vectors at every time point.

The results showed that the performance gradually increased as more data included the latter time. There was a performance difference of approximately 0.12 (F1 score, with text vector) when the whole-period data were included compared to when the last week data were excluded. The classification performance was generally better if text was included at almost all time points. When text data were included, the predicted performance increased by a mean of approximately 0.085 (F1 score), at every time point.

Table 3. Performance comparison.

Data period and text inclusion	Accuracy	F1 score	AUROC[a]
Excluding the last week			
Without text vector	0.70	0.68	0.70
With text vector	0.78	0.77	0.77
Excluding the last 5 days			
Without text vector	0.70	0.68	0.71
With text vector	0.79	0.78	0.78
Excluding the last 3 days			
Without text vector	0.71	0.70	0.72
With text vector	0.80	0.79	0.80
Including full duration data			
Without text vector	0.83	0.83	0.82
With text vector	0.89	0.89	0.89

[a] AUROC: area under the receiver operating characteristic curve.

Model Interpretation

According to the results, the contribution of the *steps per day* variable, which denotes that the number of daily steps collected automatically, not those input by the user, was larger than the contributions of the other variables (0.1085). This was followed by *afternoon snack cal* (calorie intake from snacks during the afternoon) at 0.0999 and *receive sent ratio* (ratio of the received messages to the sent messages) at 0.0967.

Among the top 20 variables, there were 12 variables related to text messages, 6 variables related to meals, 1 variable related to walking, and 1 variable related to weight. Among the 9 topic vectors, *topic bad habit max*, which corresponds to poor lifestyle patterns (such as drinking alcohol, overeating, and late-night eating) showed the highest contribution (0.0875) (Figure 3).

To verify the contribution of each variable to the churn prediction model over time, the contribution of the variables corresponding to each time over 112 days (for the 16-week program) is expressed in line plots in Figures 4 and 5 (using force_plot [28]). Overall, the contributions of the variables appeared larger as they approached the later time points of churned and retained users.

Because the model in this study outputs the probability of churn, the graph increased to make the probability of churn high for churn users and low for retained users to decrease the probability of churn. Box plots show the contribution of the variable on each date. For example, when we check each graph of the last day's churn and retained users showing the largest contributions, *steps per day* showed the greatest value, as seen for the overall variable contribution (Figure 6). Although several topic variables appeared for both churn and retained users, such as *bad habit*–related and *coaching*-related variables, the impact of each variable on the predicted performance was in the opposite direction.

To determine if the high contributing variables also showed differences in actual values between the 2 groups (churn and retained users), we checked the actual values of the *steps per day* variable, which had the greatest contribution in Figure 3. For retained users, the graph showed the entire 112 days; however, for churn users, the service use period varied depending on the user; therefore, the data were selected at varying intervals suitable for each user's start and end dates. Comparisons of actual values showed that the churn group had maintained a relatively constant value or that the value decreased slightly from the initial date to the last date of service use (from 4774.02 to 4532.35). However, the value of the retained group had a tendency to continue to decrease by half (from 5048.37 to 2485.48) starting from a value at the time of initial service use that was similar to that of the churn group (Figure 6).

Figure 3. Average impact on model output of each variable.

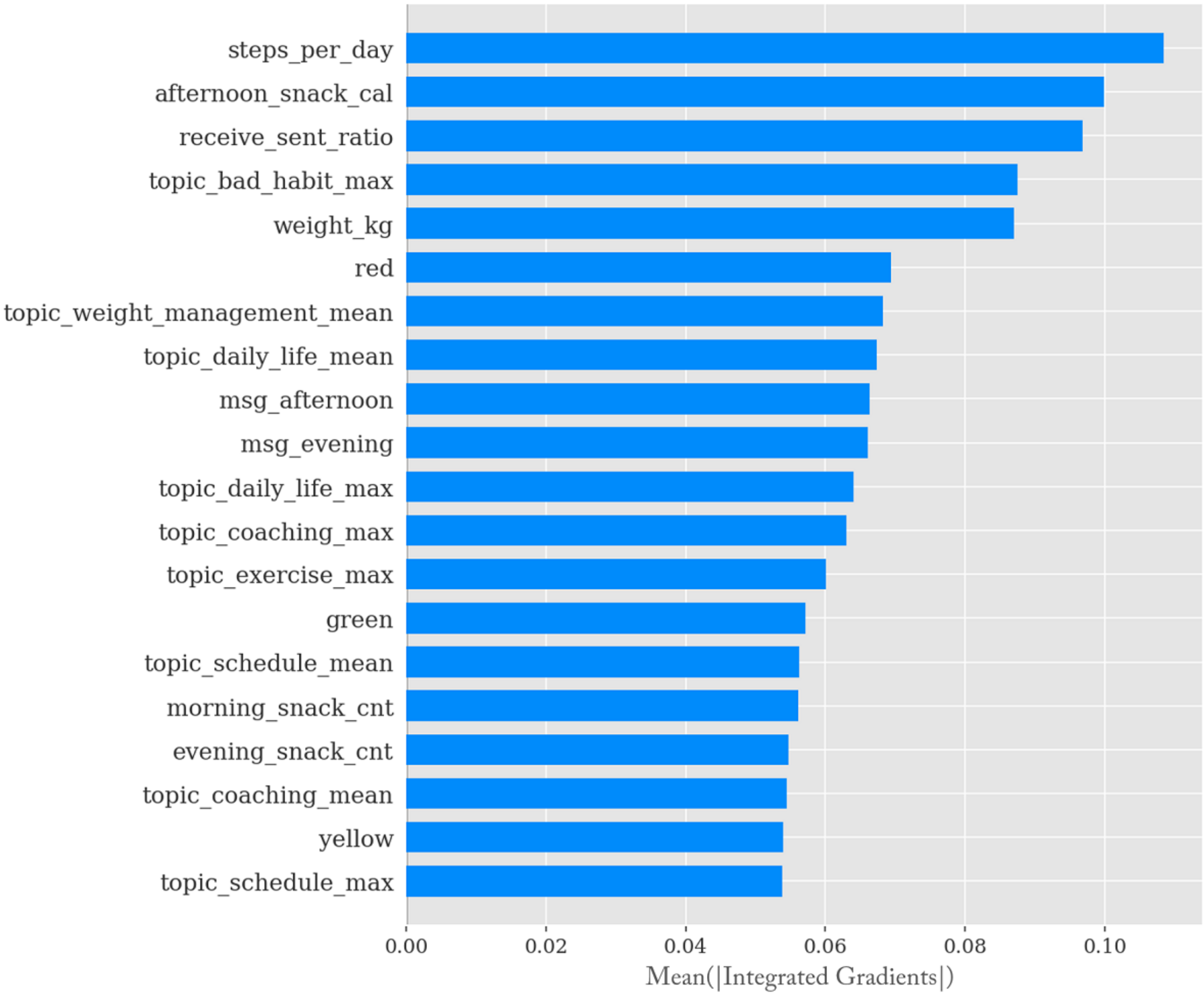


Figure 4. Churn users’ daily average contribution of variables over time (above, line plot) and contributions of each variable on the last day (below, bar plot).

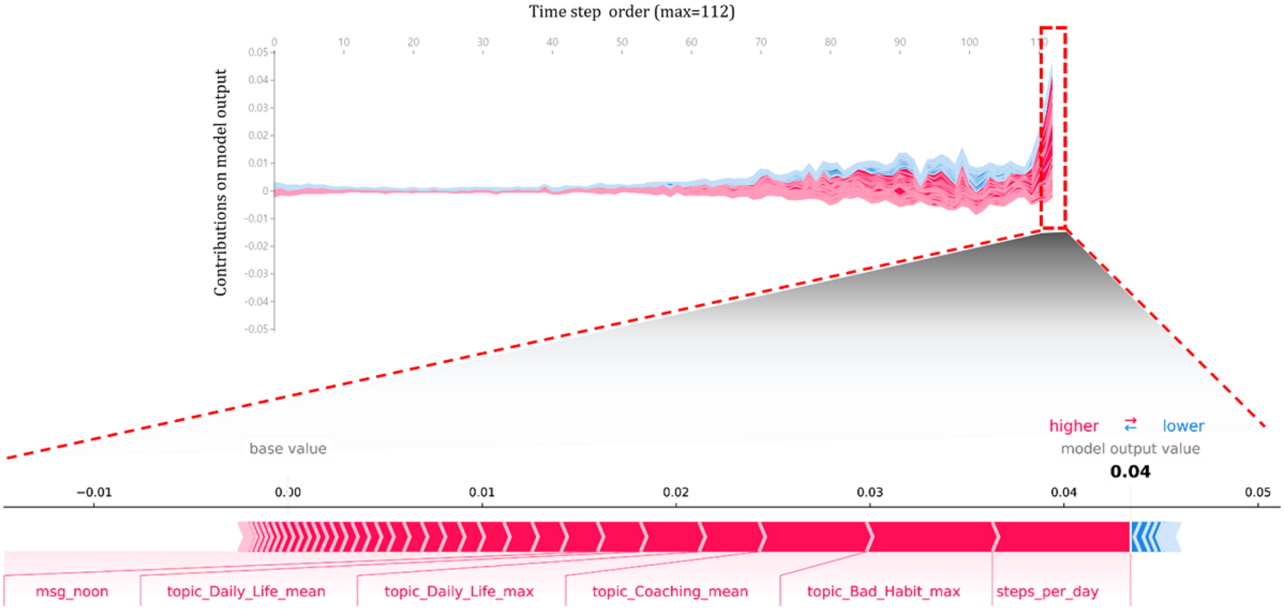


Figure 5. Retained users' daily average contribution of variables over time (above, line plot) and contributions of each variable on the last day (below, bar plot).

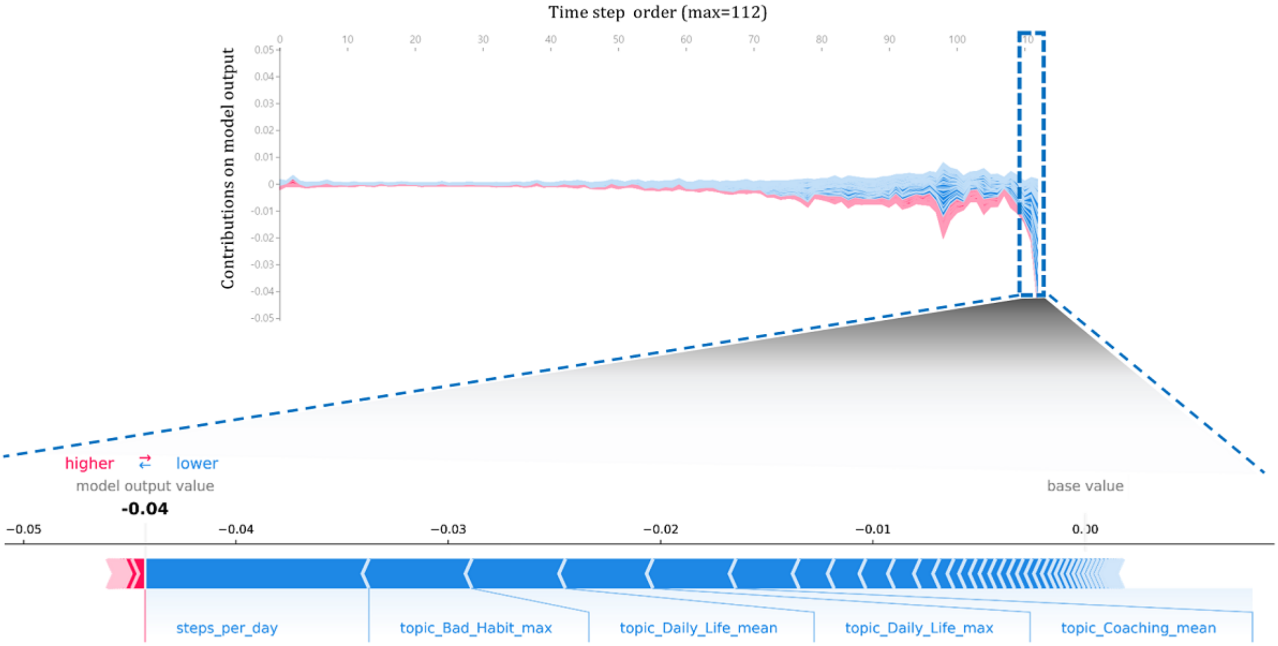
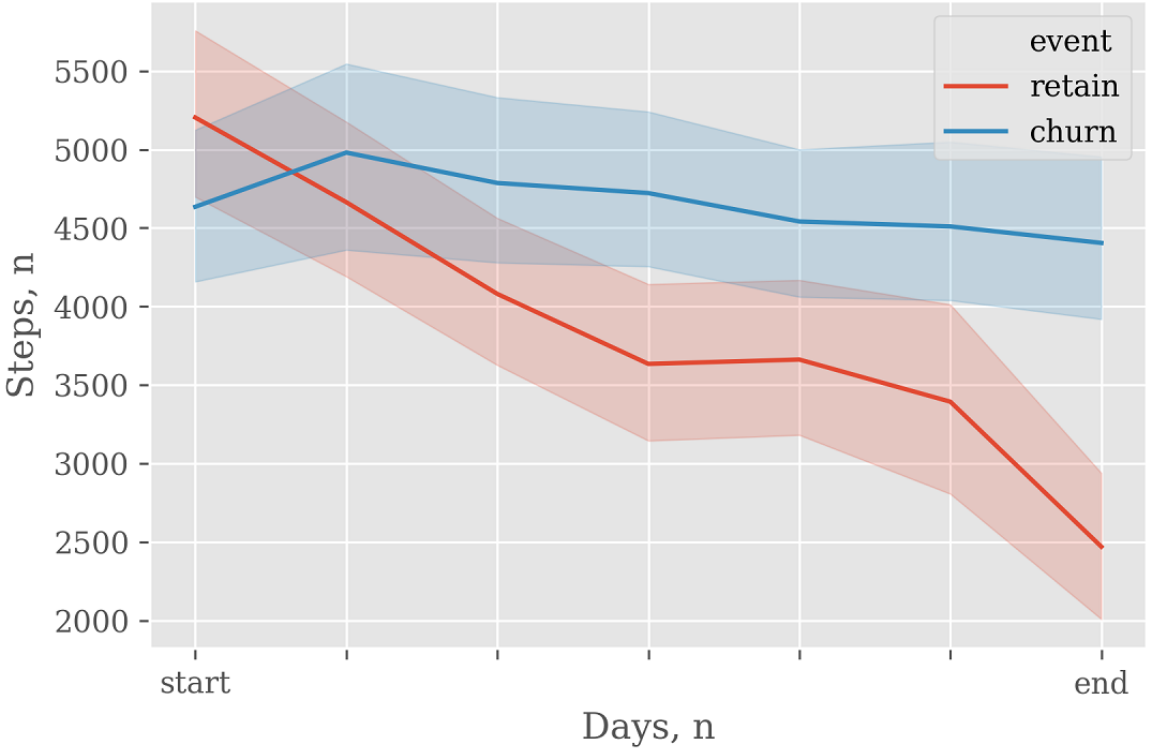


Figure 6. Changes in the mean number of steps_per_day from start to end.



Discussion

Principal Findings

First, as evidenced by the model performance comparisons, for the end time-point, when there were more data, the predictive performance increased for churn and retained users. This could also be seen during the interpretation stage: the closer the end point, the higher the contribution of variables. A previous study [7] compressed the time-series data to input them in models that do not receive time-series patterns. Even in studies [29,30] using a model that could receive time-series data as input, data were lost when inputs were reduced (ie, data cut-off) in order to match the input shape of the models. In these cases, the performance of the model may be relatively poor because it does not reflect the fact that as the end of the period is approached, there is a greater difference between the churn and retained groups [31,32]. However, in our study, by including all the data for different durations of service use by each customer, our model was able to identify churn or retained patterns of the entire time period of service use, thereby improving the predictive performance. The ability to use data of different time lengths allows for the consideration of significant characteristics of the last moment and the immediate training and prediction of churn using only the data collected at that time without waiting for a certain period of time to collect data of equal lengths from all users. This makes it possible to promptly find users who are expected to churn. Therefore, companies can focus on such users and intervene, possibly improving user adherence to their mobile app. Because the steady use of mobile health care apps is closely related to improvements in the users' health [33,34], increasing adherence is crucial to enhance satisfaction with the mobile app and for practical health promotion. Therefore, churn predictions with different lengths of data are important not only for companies but also for users.

Second, including text data can also provide better predictive performance. Because of the growing amount of unstructured customer data that can be collected both inside and outside a company, companies are studying unstructured data [35,36]. In particular, textual information can serve as important information for predicting churn [10]. Nevertheless, companies still struggle with extracting meaningful information from text [37]. In our study, we were able to increase the performance of churn prediction using vectors of messages sent by users as input to the model. We created message vectors through topic modeling to understand which specific topics affected the churn prediction by checking the contribution of each topic vector. For example, the topic that had the greatest impact on churn prediction was *bad habit topic* (eg, drinking alcohol, overeating, and late-night eating, which is known to have a negative influence on weight and health [38]). In other words, mentioning and conversing about these bad habits can also affect the churn of digital health care apps. With knowledge of this beforehand, companies can take measures to encourage the user not to churn by identifying and resolving the user's problems or complaints. Therefore, adding text data not only increases the predictive performance of the churn predictive model but also enhances interpretability by providing intuitive understanding. Companies that want to identify signs of user churn in advance should consider collecting and analyzing unstructured data that directly project customers' thoughts.

Limitations

This study was conducted, using data from churn users until the day before the churn occurred and data for the entire service period of the retained users, as a proof of concept study. However, this assumption may not be appropriate in the real world because the company may not know when the customers will leave. Nevertheless, the probability distribution of the predictions for retained users tended to change the final prediction from churn to retention. In other words, the sensitivity of the churn user data are high; therefore, from the company's point of view, the possibility of missing the churn user can be reduced. However, for practical forecasting by real-world companies, further studies of variations in probability of churn over time are needed. This study was retrospective. It used past records of mobile app users to identify signals of churn and identify the churn users. Because our analysis was retrospective, there was a constraint on its effectiveness for prospective data. Therefore, it is necessary to study whether early interventions for groups expected to be churn users provide more clinical indicators.

Conclusion

We used a model with recurrent neural network architecture that used user log data and text data to determine the churn of digital health care users. Our analysis of variables is expected to help identify signs of user churn in advance and improve adherence in the field of digital health care.

Acknowledgments

This study received support from the Institute for Information & Communications Technology Promotion grant funded by the Korea government (Ministry of Science and Information and Communication Technology, grant number 13-08-0-130-133-2000–2033) and the Korea Health Technology Research & Development Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant number HI19C1015).

Conflicts of Interest

None declared.

References

1. Mozer M, Wolniewicz R, Grimes D, Johnson E, Kaushansky H. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. IEEE Trans Neural Netw 2000;11(3):690-696. [doi: 10.1109/72.846740] [Medline: 18249796]
2. Paswan A. Profit Brand: How to Increase the Profitability, Accountability and Sustainability of Brands. Journal of Product & Brand Management 2007 Jul 24;16(4):293-294. [doi: 10.1108/10610420710763994]
3. Schweitzer J, Synowiec C. The economics of eHealth and mHealth. J Health Commun 2012 May 02;17 Suppl 1(sup1):73-81. [doi: 10.1080/10810730.2011.649158] [Medline: 22548602]

4. Statista. URL: https://www.statista.com/topics/2263/mhealth/#dossierSummary__chapter3 [accessed 2020-06-20]
5. Seo D, Park YR, Lee Y, Kim JY, Park J, Lee J. The use of mobile personal health records for hemoglobin A1c regulation in patients with diabetes: retrospective observational study. J Med Internet Res 2020 Jun 02;22(6):e15372 [FREE Full text] [doi: 10.2196/15372] [Medline: 32484447]
6. Lee J, Yoon W, Cho S, Cho Y. A strategic consideration on chronic disease prevention and management through mHealth utilization. KJPH 2016 Mar 31;53(1):7. [doi: 10.17262/kjph.2016.03.53.1.7]
7. Lee D, Shin M. A dynamic study on the analysis of customer departure using recurrent neural network. 2017 Presented at: The Korea Society of Management information Systems Annual Conference on Management; 2017; Seoul p. 290-294.
8. Yang Y, Yang Q, Lu W, Pan J, Pan R, Lu C, editors. Preprocessing time series data for classification with application to CRM. 2005 Presented at: Australasian Joint Conference on Artificial Intelligence; 2005; Sydney. [doi: 10.1007/11589990_16]
9. Schatzmann A, Heitz C, Münch T. Churn prediction based on text mining and CRM data analysis. 2014 Presented at: 13th International Science-to-Business Marketing Conference:«Cross Organizational Value Creation»; 2-4 June; Winterthur.
10. De Caigny A, Coussement K, De Bock KW, Lessmann S. Incorporating textual information in customer churn prediction models based on a convolutional neural network. International Journal of Forecasting 2020 Oct;36(4):1563-1578. [doi: 10.1016/j.ijforecast.2019.03.029]
11. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36(1):393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
12. Kim Y, Oh B, Shin H. Effect of mHealth with offline antiobesity treatment in a community-based weight management program: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 21;8(1):e13273 [FREE Full text] [doi: 10.2196/13273] [Medline: 31961335]
13. Burez J, Van den Poel D. Handling class imbalance in customer churn prediction. Expert Systems with Applications 2009 Apr;36(3):4626-4636. [doi: 10.1016/j.eswa.2008.05.027]
14. Cho JH. The investigation of factors of determining continuous use of health apps on smartphones application of extended technology acceptance model. Journal of Public Relations Research 2014 Feb;18(1):212-241. [doi: 10.15814/jpr.2014.18.1.212]
15. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
16. Hochreiter S, Bengio Y, Frasconi P, Schmidhuber J. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In: A Field Guide to Dynamical Recurrent Neural Networks. New York City: IEEE Press; 2001.
17. Lipton Z, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning. arXiv. 2015. URL: https://arxiv.org/pdf/1506.00019.pdf [accessed 2020-12-18]
18. Zhu S, Chollet F. Developer guides.: Keras; 2019. URL: https://keras.io/guides/understanding_masking_and_padding [accessed 2020-06-20]
19. Steyvers M, Griffiths T. Probabilistic topic models. In: Landauer TK, McNamara DS, Dennis S, Kintsch W, editors. Handbook of Latent Semantic Analysis. New York, NY: Psychology Press; 2007.
20. Blei D, Ng A, Jordan M. Latent dirichlet allocation. Journal of Machine Learning Research 2003;3:993-1022 [FREE Full text]
21. Kim MK, Lee Y, Han CH. A Study on Priority Determination Using Topic-Modeling and AHP. 2018 Presented at: Korean Society for Industrial Management Systems Conference; October 31; Seoul p. 10.
22. Mimno D, Wallach H, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. 2011 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing; July; Edinburgh URL: https://www.aclweb.org/anthology/D11-1024.pdf
23. Samek W, Müller KR. Towards explainable artificial intelligence. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller KR, editors. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. New York City: Springer; 2019:5-22.
24. Ancona M, Ceolini E, Öztireli C, Gross M. Gradient-based attribution methods. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller KR, editors. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. New York City: Springer; 2019:169-191.
25. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv. 2013. URL: https://arxiv.org/pdf/1312.6034.pdf [accessed 2020-12-18]
26. Alber M. Software and application patterns for explanation methods. In: Samek W, Montevon G, Vedaldi A, Hansen LK, Müller KR, editors. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. New York City: Springer; 2019:399-433.
27. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. arXiv. 2017. URL: https://arxiv.org/pdf/1703.01365.pdf [accessed 2020-12-18]
28. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct 10;2(10):749-760 [FREE Full text] [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]

29. Martins H. Predicting user churn on streaming services using recurrent neural networks. Digitala Vetenskapliga Arkivet. 2017. URL: https://www.diva-portal.org/smash/get/diva2:1153731/FULLTEXT01.pdf [accessed 2020-12-18]
30. Hu J, Zhuang Y, Yang J, Lei L, Huang M, Zhu R. A recurrent neural network based approach for customer churn prediction in telecommunication sector. 2018 Presented at: IEEE International Conference on Big Data; December 10-13; Seattle, WA. [doi: 10.1109/bigdata.2018.8622094]
31. Huang Y, Zhu F, Yuan M, Deng K, Li Y, Ni B. Telco churn prediction with big data. 2015 Presented at: Proceedings of the ACM SIGMOD international conference on management of data; May 31-June 4; Melbourne. [doi: 10.1145/2723372.2742794]
32. Mena C, De CA, Coussement K, De BK, Lessmann S. Churn prediction with sequential data and deep neural networks a comparative analysis. arXiv. 2019. URL: https://arxiv.org/pdf/1909.11114.pdf [accessed 2020-12-18]
33. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
34. Yi JY, Kim Y, Cho Y, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
35. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 2015 Apr;35(2):137-144. [doi: 10.1016/j.ijinfomgt.2014.10.007]
36. Sheng J, Amankwah-Amoah J, Wang X. A multidisciplinary perspective of big data in management research. International Journal of Production Economics 2017 Sep;191:97-112. [doi: 10.1016/j.ijpe.2017.06.006]
37. Kumar BS, Ravi V. A survey of the applications of text mining in financial domain. Knowledge-Based Systems 2016 Dec;114:128-147. [doi: 10.1016/j.knosys.2016.10.003]
38. Elfhag K, Rossner S. Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. Obesity Reviews 2005 Feb;6(1):67-85. [doi: 10.1111/j.1467-789x.2005.00170.x]

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by N Mohammad Gholi Mezerji, F Muehlensiepen; comments to author 14.09.20; revised version received 02.11.20; accepted 05.11.20; published 06.01.21.

Please cite as:
Kwon H, Kim HH, An J, Lee JH, Park YR
Lifelog Data-Based Prediction Model of Digital Health Care App Customer Churn: Retrospective Observational Study
J Med Internet Res 2021;23(1):e22184
URL: https://www.jmir.org/2021/1/e22184
doi:10.2196/22184
PMID:33404511

Original Paper

Factors Associated With Dropout of Participants in an App-Based Child Injury Prevention Study: Secondary Data Analysis of a Cluster Randomized Controlled Trial

Jie Li[1], BA; Peishan Ning[1], PhD; Peixia Cheng[1], PhD; David C Schwebel[2], PhD; Yang Yang[3], PhD; Xiang Wei[1], MD; Jieyi He[1], BA; Wanhui Wang[1], BA; Ruotong Li[1], BA; Guoqing Hu[1], PhD

[1]Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

[2]Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, United States

[3]Department of Biostatistics, College of Public Health and Health Professions, Emerging Pathogen Institute, University of Florida, Gainesville, FL, United States

Corresponding Author:
Guoqing Hu, PhD
Department of Epidemiology and Health Statistics
Xiangya School of Public Health
Central South University
110 Xiangya Road
Changsha, 410078
China
Phone: 86 731 84805414
Email: huguoqing009@gmail.com

Abstract

Background: Mobile health (mHealth) interventions offer great potential to reach large populations and improve public health. However, high attrition rates threaten evaluation and implementation of mHealth intervention studies.

Objective: We explored factors associated with attrition of study participants in an mHealth randomized controlled trial (RCT) evaluating an intervention to reduce unintentional child injury risk in China.

Methods: The cluster RCT compared two groups of an app-based intervention for caregivers of 3-6–year-old children (Bao Hu San). The intervention group received unintentional child injury and parenting education, whereas only parenting education was implemented in the control group. The trial included 2920 study participants in Changsha, China, and lasted 6 months. Data on participant engagement (using the app) were collected electronically throughout the 6-month period. Associations between participant attrition and demographic characteristics, and between attrition and intervention engagement were tested and quantified separately for the intervention and control groups using the adjusted odds ratio (aOR) based on generalized linear mixed models.

Results: In total, 2920 caregivers from 20 eligible preschools participated, with 1510 in the intervention group and 1410 in the control group. The 6-month attrition rate differed significantly between the two groups ($P<.001$), at 28.9% (437/1510) in the intervention group and 35.7% (503/1410) in the control group. For the intervention group, the only significant predictor of attrition risk was participants who learned fewer knowledge segments (aOR 2.69, 95% CI 1.19-6.09). For the control group, significant predictors of attrition risk were lower monthly login frequency (aOR 1.48, 95% CI 1.00-2.18), learning fewer knowledge segments (aOR 1.70, 95% CI 1.02-2.81), and shorter learning durations during app engagement (aOR 2.39, 95% CI 1.11-5.15). Demographic characteristics were unrelated to attrition.

Conclusions: Engagement in the app intervention was associated with participant attrition. Researchers and practitioners should consider how to best engage participants in app-based interventions to reduce attrition.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-IOR-17010438; http://www.chictr.org.cn/showproj.aspx?proj=17376

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-018-5790-1

(J Med Internet Res 2021;23(1):e21636) doi:10.2196/21636

KEYWORDS

app-based intervention; unintentional injury; attrition; influencing factors

Introduction

Owing to recognized advantages such as cost-effective dissemination, real-time data collection and feedback, reduced burden, flexible customization, self-monitoring capacity, and visually attractive multimedia presentation [1-3], mobile health (mHealth) technology has become increasingly popular in health intervention research and practice over the past decade. A wide range of mHealth interventions have been developed to prevent diseases and injuries, increase the compliance to recommended health interventions, and offer remote access to health services [4-10].

Despite these advantages, mHealth interventions suffer from the challenge of high dropout attrition rates compared to studies adopting traditional interventions. For example, a recent 6-month large intervention study for smoking cessation reported an attrition rate of 57% in the intervention group using an app and of only 52% in a control group using a self-help booklet as usual education [11]. Several app-based intervention studies reported high attrition rates, with estimates ranging from 38% to 84% [12-16].

When attrition is high, occurs unequally between intervention and control groups, or occurs in a nonrandom way, it threatens the validity of evaluation studies [17,18]. When dropout attrition is high in public health practice, the efficacy of an intervention is lower than desired. Therefore, researchers must prioritize exploration and understanding of the factors associated with dropout attrition of study participants. Increased understanding would lead to feasible approaches to prevent and reduce dropout attrition.

Previous research examining predictors of dropout attrition from app-based intervention studies in reducing alcohol intake [19], improving health-related behavior [20], and weight loss [21] focused primarily on the demographic characteristics of individuals who are more likely to fail to complete follow-up surveys and therefore are not retained in the research study. For example, in an evaluation of the efficacy of an app-based child burn prevention program, Burgess et al [22] reported a higher proportion of university degree holders among participants who remained in the study (28.7%, 70/244) than those who were lost to follow up (16.5%, 42/254). Despite the value of this line of research, it is equally important to explore relationships between attrition and other indicators such as engagement with the app-based intervention. Such analyses may offer new clues to reducing attrition in app-based intervention studies.

Therefore, the aim of this study was to examine associations both between attrition and demographic characteristics and between attrition and participants’ engagement in the intervention. We hypothesized that a more active and attractive type of intervention would result in greater engagement and thus a lower attrition rate in the intervention arm compared with the control arm. Replicating Burgess et al [22] and others, we also hypothesized that greater education might be related to higher retention in the intervention study. We used data collected from a cluster randomized controlled trial (RCT) examining an app intervention to prevent unintentional child injury in China to test our hypotheses.

Methods

Design

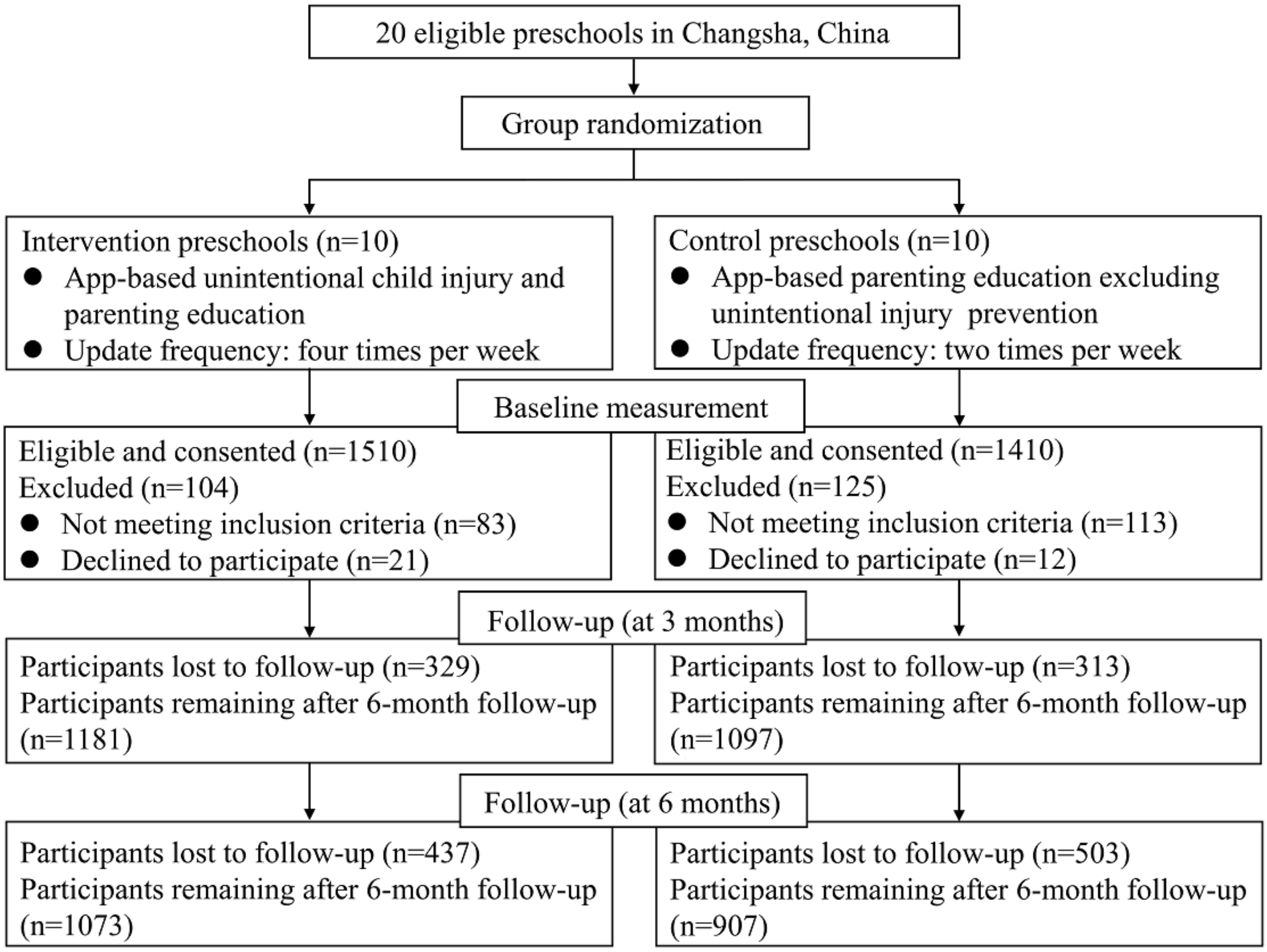
This research comprises secondary analyses of data from a published single-blinded cluster RCT assessing the effectiveness of an app-based intervention for caregivers to prevent unintentional injury among Chinese preschoolers [8]. The trial recruited 2920 caregivers of 3-6–year-old children from 20 eligible preschools using cluster random sampling. The randomization was performed at the school level and was stratified by type of preschool (public vs private), yielding five public and five private schools in both the intervention group (implementing unintentional child injury and parenting education) and the control group (implementing parenting education only).

Both groups engaged in a 6-month intervention delivered through a smartphone app. The intervention group received a more active and attractive intervention than the control group. Specifically, the control group received essays, games, comics, and videos twice a week, and users conducted thematic discussion activity once a month to learn about common children’s diseases and to practice parenting skills outside unintentional injury prevention. In contrast, the intervention group received all content that the control group received, but the participants were also exposed to additional similar essays, games, comics, and videos twice a week, and engaged in a second set of thematic discussions once a month to learn about parenting skills related to unintentional child injury prevention. The intervention group also had access to an app-based portal that supported communication between users and professionals concerning injury prevention knowledge and skills.

Data were collected through three caregiver surveys that were completed at baseline in December 2017 and then two follow-up surveys in March and June 2018, which corresponded to the 3rd and 6th month after the initiation of the interventions in both groups. To encourage adherence in using the app, reminders about newly released knowledge segments appeared on the scroll screen when users opened the app. Users who did not read the content in a timely manner were reminded again. We also provided financial incentives to participants after completing study of the knowledge segments and asked the preschool teachers to remind caregivers regularly to use the app. In addition, an automatic message to remind participants about using the app was sent if the user failed to log in for 1 month. To encourage compliance in completing the study surveys, we provided financial incentives and sent two automatic reminder messages to users who did not complete the follow-up surveys 3 days after they were released.

Figure 1 illustrates the study design. The protocol was approved by the Ethics Committee of Xiangya School of Public Health, Central South University, Changsha, China (approval number: XYGW-2017-02). Full details of the RCT evaluation are published elsewhere [23].

Figure 1. Flow diagram of the app-based intervention study.



Outcome Measure

Attrition from the study was defined as a study participant who failed to complete both follow-up surveys at the 3rd and 6th months after initiation of the interventions.

Independent Variables

Based on our study hypotheses and previous studies [17,22], we considered demographic factors and engagement with the app-based intervention as potential factors that would predict dropout attrition of study participants. We also considered previous training experience on unintentional injury prevention, as we hypothesized that participants might engage in the intervention or study only if they valued parenting education and if they were receiving new information that they had not learned in the past.

Data on demographic characteristics and previous training experience were obtained via a baseline survey. Caregiver demographic factors included sex, age, level of education, and monthly household income per capita. Caregiver's age was divided into three approximately equally sized groups (less than 31 years, 31 to 34 years, and 35 years and above). Level of education was classified into three categories: junior high school and below, high school (including technical secondary school), and junior college and above. Monthly household income per capita was divided into two levels: below and above 3500 yuan (approximately US $540), according to the average income of Changsha residents in 2015 [24].

Data on engagement with the intervention were collected using the app's backend system across full implementation of the trial; that is, user engagement information was collected in an automated fashion when users logged into the app. Available data included the webpages on the app that were visited, and on what occasions, how frequently, and for how long. This kind of information offers insight into engagement by participants based on how much time they spend with the app, and with what app features.

For this study, we collected app engagement data using four indicators: monthly login frequency, single login duration, knowledge segments learned per login, and single learning duration. Monthly login frequency was defined as the average number of times each participant used the app per month before quitting the study. Quitting was defined as the date of the last login recorded in the app backend system. Single login duration was the average online duration between opening the app and exiting it in the same session. Knowledge segments learned per login was the average number of short written statements with pictures, cartoon episodes, video recommendations, and interactive games learned in each login. Finally, single learning duration was defined as the average time spent on reading and learning individual knowledge segments during each login. To facilitate interpretation of the results, we divided all participants into three equal groups for each indicator according to the percentiles of $P_{33.4}$ and $P_{66.7}$ ($<P_{33.4}$, $P_{33.4}$ to $P_{66.7}$, and $>P_{66.7}$).

Statistical Analysis

Attribution rates and 95% CIs were estimated based on binomial distributions. The χ^2 test was used to examine the difference in attrition rates between the intervention and control education groups. Differences in app intervention engagement measures between the two groups were evaluated using the Kruskal-Wallis *H* test. Generalized linear mixed models were used to test the associations between attrition and all independent variables. The intraclass correlation coefficient was calculated to quantify the clustering of study participants at the preschool level. Adjusted odds ratios (aORs) were calculated to quantify the size of associations after adjusting for other independent variables. All analyses were performed using SAS 9.2 (SAS Institute). Statistical significance was based on two-sided tests at the level of .05.

Results

Sample Characteristics

In total, 2920 caregivers from 20 eligible preschools in Changsha, China participated, with 1510 caregivers assigned to the intervention group and 1410 assigned to the control group (Table 1). Participants in the intervention group were slightly more likely to be male, to have a higher monthly household income per capita, and to have received injury prevention education in the past 3 months. There also were differences in engagement: participants in the intervention group were more likely to engage in the website based on all four metrics than those in the control group (P<.001).

Table 1. Demographic and engagement characteristics of study participants.

Characteristic	Intervention (n=1510)	Control (n=1410)	*P* value
Age (years), mean (SD)	32.7 (5.0)	33.1 (5.7)	.06
Sex, n (%)			
Male	490 (32.5)	364 (25.8)	<.001
Female	1020 (67.5)	1046 (74.2)	
Educational level, n (%)			
Junior high school and below	1017 (67.3)	951 (67.5)	.98
High school	386 (25.6)	357 (25.3)	
College and above	107 (7.1)	102 (7.2)	
Household income per capita per month (US $), n (%)			
<540	367 (24.3)	272 (19.3)	.001
≥540	1143 (75.7)	1138 (80.7)	
Injury prevention education in past 3 months, n (%)			
Yes	918 (60.8)	708 (50.2)	<.001
No	592 (39.2)	702 (49.8)	
Monthly login frequency, median (IQR)	3.9 (5.5)	3.0 (5.5)	<.001
Single login duration (seconds), median (IQR)	204.9 (221.8)	176.8 (198.7)	<.001
Knowledge segments learned per login, median (IQR)	0.7 (1.7)	0.5 (1.2)	<.001
Single learning duration (seconds), median (IQR)	61.9 (162.8)	44.9 (131.0)	<.001

Engagement Indicators Between the Intervention and Control Groups

The intervention group had significantly higher engagement than the control group based on all four engagement indicators. This was true both for participants who completed the study and for those who failed to complete the study (Table 2).

Table 2. Differences in app engagement indicators between the intervention and control groups within 6 months.

Variable	Completers (n=1980)			Lost to follow up (n=940)		
	Intervention, median (IQR)	Control, median (IQR)	*P* value	Intervention, median (IQR)	Control, median (IQR)	*P* value
Monthly login frequency (N)	3.9 (6.0)	3.5 (6.0)	.03	3.9 (4.6)	2.0 (4.0)	<.001
Single login duration (seconds)	222.2 (232.6)	195.0 (214.2)	<.001	156.3 (174.2)	142.6 (192.9)	<.001
Knowledge segments learned per login (N)	1.0 (1.7)	0.8 (1.3)	<.001	0.2 (0.9)	0.1 (0.7)	.02
Single learning duration (seconds)	86.0 (172.2)	72.7 (163.7)	.01	16.0 (82.8)	7.4 (54.3)	.009

Attrition Rate

The 6-month attrition rate for the intervention group was 28.9% (437/1510; 95% CI 26.6%-31.2%), whereas that for the control group was 35.7% (503/1410; 95% CI: 33.2%-38.2%). The attrition was significantly higher in the control group than in the intervention group (OR 1.36, 95% CI 1.17-1.59) (Multimedia Appendix 1).

Influencing Factors

Given the distinct attrition rates across the two groups, we performed multivariable analysis separately for each group. As shown in Table 3, the attrition rates of caregivers in the intervention group with “sometimes” monthly login frequency was higher than that of those with “often” monthly login frequency (*P*=.03). The attrition rates of caregivers with short and average single login durations were significantly higher than those for caregivers with long single login durations (*P*<.001 and *P*=.03, respectively). The attrition rates of caregivers with few and average number of knowledge segments learned per login were significantly higher than that for caregivers completing many knowledge segments learned per login (*P*<.001 and *P*=.003, respectively). Caregivers with short and average single learning durations had higher attrition rates than those with a long single learning duration (*P*<.001 and *P*=.002, respectively). The multivariable analysis showed that caregivers who learned fewer knowledge segments per login had a substantially higher attrition risk than those who learned more knowledge segments per login after adjusting for sex, age group, level of education, monthly household income per capita, receiving injury prevention education in the past 3 months or not, and the three other intervention engagement measures (Table 3).

As shown in Table 4, the attrition rate of caregivers in the control group with “seldom” monthly login frequency was higher than that for those with “often” monthly login frequency (*P*=.003). The attrition rates of caregivers with short and average single login durations were higher than that of caregivers with long single login durations (*P*<.001 and *P*=.009, respectively). The attrition rates of caregivers with few and average knowledge segments learned per login were much higher than that for caregivers with many knowledge segments learned per login (*P*<.001 and *P*=.001, respectively). The attrition rates of caregivers with short and average single learning durations were higher than that for caregivers with long single learning durations (*P*<.001 and *P*=.002, respectively). The multivariable analysis for attrition yielded three significant predictors of attrition. Caregivers with a lower monthly login frequency, who learned fewer knowledge segments per login, and who had shorter single learning durations had significantly higher attrition rates than those using the app more often per month (seldom vs often), learning more knowledge segments per login (average vs many), and having a longer single learning duration (short vs long) (Table 4).

Table 3. Six-month attrition rates of caregivers of preschoolers in the intervention group (N=1510).

Variable	Attrition rate (%) (95% CI)	OR[a] (95% CI)	aOR[b] (95% CI)
Total[c]	28.9 (26.6-31.2)	N/A[d]	N/A
Sex			
Male	24.3 (20.5-28.1)	Reference	Reference
Female	31.2 (28.4-34.0)	1.14 (0.85-1.54)	1.24 (0.90-1.70)
Age group (years)			
<31	34.6 (30.4-38.8)	1.43 (1.00-2.05)	1.32 (0.90-1.94)
31-34	28.4 (24.8-32.0)	1.32 (0.92-1.91)	1.36 (0.93-1.99)
≥35	23.0 (19.0-27.0)	Reference	Reference
Level of education			
Junior high school and below	40.2 (30.9-49.5)	1.28 (0.76-2.17)	1.23 (0.69-2.22)
High school	32.4 (27.7-37.1)	0.97 (0.70-1.35)	0.98 (0.68-1.40)
College and above	26.5 (23.8-29.2)	Reference	Reference
Monthly household income per capita (US $)			
<540	31.9 (27.1-36.7)	0.96 (0.70-1.33)	0.90 (0.63-1.30)
≥540	28.0 (25.4-30.6)	Reference	Reference
Received injury prevention education in the past 3 months			
Yes	26.3 (23.5-29.1)	Reference	Reference
No	33.1 (29.3-36.9)	1.20 (0.90-1.60)	1.09 (0.80-1.48)
Monthly login frequency[e]			
Seldom ($<P_{33.4}$)	31.7 (27.4-36.0)	1.24 (0.88-1.74)	0.88 (0.60-1.29)
Sometimes ($P_{33.4}$-$P_{66.7}$)	30.5 (26.6-34.4)	1.46 (1.05-2.01)	1.26 (0.89-1.78)
Often ($>P_{66.7}$)	25.2 (21.6-28.8)	Reference	Reference
Single login duration[e]			
Short ($<P_{33.4}$)	40.4 (35.8-45.0)	2.68 (1.91-3.77)	1.26 (0.83-1.90)
Average ($P_{33.4}$-$P_{66.7}$)	27.9 (24.0-31.8)	1.49 (1.06-2.09)	1.01 (0.69-1.48)
Long ($>P_{66.7}$)	20.7 (17.3-24.1)	Reference	Reference
Knowledge segments learned per login[e]			
Few ($<P_{33.4}$)	48.9 (44.3-53.5)	4.76 (3.34-6.77)	2.69 (1.19-6.09)
Average ($P_{33.4}$-$P_{66.7}$)	25.9 (22.0-29.8)	1.86 (1.30-2.66)	1.38 (0.83-2.30)
Many ($>P_{66.7}$)	15.9 (12.9-18.9)	Reference	Reference
Single learning duration[e]			
Short ($<P_{33.4}$)	48.1 (43.5-52.7)	4.67 (3.26-6.67)	1.76 (0.78-3.95)
Average ($P_{33.4}$-$P_{66.7}$)	26.1 (22.3-29.9)	1.92 (1.34-2.75)	1.42 (0.88-2.31)
Long ($>P_{66.7}$)	15.9 (12.9-18.9)	Reference	Reference

[a] OR: odds ratio.

[b] aOR: adjusted odds ratio.

[c] The intraclass correlation coefficient was 0.15 for level two (preschool). Tests for multicollinearity indicated a low level of multicollinearity (tolerance>0.10 and variance inflation factor<5 for all predictors). The Hosmer-Lemeshow goodness-of-fit suggested that overall model fit was acceptable (*P*=.75).

[d] N/A: not applicable.

[e] These variables were equally divided into three groups based on the $P_{33.4}$ and $P_{66.7}$ percentiles.

Table 4. Six-month attrition rates of caregivers of preschoolers in the control group (N=1410).

Variable	Attrition rate (%) (95% CI)	OR[a] (95% CI)	aOR[b] (95% CI)
Total[c]	35.7 (33.2-38.2)	N/A[d]	N/A
Sex			
Male	33.2 (28.4-38.0)	Reference	Reference
Female	36.5 (33.6-39.4)	1.04 (0.76-1.42)	1.05 (0.76-1.47)
Age group (years)			
<31	38.4 (33.7-43.1)	1.03 (0.72-1.48)	0.99 (0.68-1.45)
31-34	38.0 (34.0-42.0)	1.25 (0.90-1.74)	1.15 (0.81-1.63)
≥35	30.2 (25.9-34.5)	Reference	Reference
Education level			
Junior high school and below	47.1 (37.4-56.8)	1.13 (0.68-1.87)	1.02 (0.58-1.80)
High school	40.1 (35.0-45.2)	0.94 (0.69-1.30)	0.88 (0.63-1.24)
College and above	32.8 (29.8-35.8)	Reference	Reference
Monthly household income per capita (US $)			
<540	38.6 (32.8-44.4)	1.02 (0.73-1.42)	1.04 (0.71-1.51)
≥540	35.0 (32.2-37.8)	Reference	Reference
Received injury prevention education in past 3 months			
Yes	30.6 (27.2-34.0)	Reference	Reference
No	40.7 (37.1-44.3)	1.20 (0.91-1.58)	1.19 (0.89-1.59)
Monthly login frequency[e]			
Seldom ($<P_{33.4}$)	46.2 (42.0-50.4)	1.77 (1.27-2.48)	1.48 (1.00-2.18)
Sometimes ($P_{33.4}$-$P_{66.7}$)	30.5 (26.1-34.9)	1.03 (0.72-1.48)	1.10 (0.75-1.61)
Often ($>P_{66.7}$)	27.4 (23.2-31.6)	Reference	Reference
Single login duration[e]			
Short ($<P_{33.4}$)	44.7 (40.5-48.9)	2.15 (1.53-3.02)	0.98 (0.65-1.48)
Average ($P_{33.4}$-$P_{66.7}$)	34.3 (30.0-38.6)	1.68 (1.18-2.39)	1.21 (0.82-1.80)
Long ($>P_{66.7}$)	25.9 (21.7-30.1)	Reference	Reference
Knowledge segments learned per login[e]			
Few ($<P_{33.4}$)	53.6 (49.3-57.9)	4.43 (3.04-6.46)	1.92 (0.88-4.20)
Average ($P_{33.4}$-$P_{66.7}$)	31.5 (27.4-35.6)	2.15 (1.46-3.15)	1.70 (1.02-2.81)
Many ($>P_{66.7}$)	17.5 (13.7-21.3)	Reference	Reference
Single learning duration[e]			
Short ($<P_{33.4}$)	54.2 (49.9-58.5)	4.29 (2.98-6.19)	2.39 (1.11-5.15)
Average ($P_{33.4}$-$P_{66.7}$)	30.4 (26.3-34.5)	1.98 (1.36-2.89)	1.45 (0.89-2.35)
Long ($>P_{66.7}$)	18.5 (14.8-22.2)	Reference	Reference

[a] OR: odds ratio.

[b] aOR: adjusted odds ratio.

[c] The intraclass correlation coefficient was 0.15 for level two (preschool); tests for multicollinearity indicated a low level of multicollinearity (tolerance>0.10 and variance inflation factor<5 for all predictors); the Hosmer-Lemeshow goodness-of-fit suggested that overall model fit was acceptable (*P*=.85).

[d] N/A: not applicable.

[e] These variables were equally divided into three groups based on the $P_{33.4}$ and $P_{66.7}$ percentiles.

Discussion

Principal Findings

The 6-month attrition rate of study participants was 28.9% in the intervention group and was 35.7% in the control group. As hypothesized, higher attrition risks were associated with low app intervention engagement in both the intervention and control groups after controlling for demographic variables and previous training experience on unintentional injury prevention. Compared to the intervention group, the control group had a higher attrition rate, which was primarily explained by lower app intervention engagement. Demographic factors seemed to not be associated with attrition in either group.

Interpretation of Findings

Attrition in our study was quite similar to that reported at 6-month follow up for an educational app on car seat use [25] (34.3%, 387/1129), but was significantly lower than the rate of 51% (254/498) reported in an app-based study of scald burn prevention [22]. A wide range of factors likely explain participant attrition from engagement in an app-based intervention, including design of the app, assessment measures for participant compliance, individual characteristics of participants, and cultural differences across countries.

Previous studies [13,22,26] generally report that female, younger, and less educated individuals have a higher likelihood of attrition. We could not confirm these findings in either the univariable or the multivariable analyses. One possible explanation for this result is the fact that we conducted our study in China, a middle-income country with a collectivist culture that differs from other countries and regions where the previous studies were conducted, such as the United States and Australia. Another possible explanation is related to our multivariable findings, which indicate that participant engagement was a strong predictor of participant attrition after controlling for participant demographics. In both groups, engagement in the intervention was significantly associated with attrition. These results support the hypothesis that a high proportion of participants lost to follow up in app-based interventions may simply be a result of nonusage of the app [17]. If participants perceive the app as helpful and beneficial, they continue to use it and then engage in research surveys.

We found that the attrition rate in the intervention group was lower than that in the control group, a finding we attribute to the less engaging aspects of the parenting education app for the control group. Specifically, the control group received new material twice a week and thematic discussions once a month, while the intervention group received new material four times a week and engaged in thematic discussions twice a month. The intervention group also had access to experts to ask questions and receive consultation, a feature not provided to the control group. These data concord with the findings of Kelders et al [27] who reviewed 101 web-based intervention studies and found that increased interaction with a counselor, more frequent intended usage, more frequent updates, and more extensive employment of dialogue support better adherence.

Implications

Our findings have two major implications. First, they emphasize the unignorable impact of high dropout attrition in mHealth intervention studies. High attrition must be evaluated strictly before drawing conclusions about the efficacy of an intervention, and research findings should be interpreted cautiously when attrition rates are high, vary substantially across comparison groups, or when missing values (dropouts) occur in a nonrandom fashion. Second, the results underscore the need for careful design and implementation of mHealth interventions. The comparatively high attrition rate among participants with less frequent and engaging exposure to educational materials highlights the need for mHealth interventions to be designed and implemented with appealing features. They must attract users and increase engagement in the intervention.

Study Limitations

This study was primarily limited by the RCT design. We do not have data concerning engagement or attrition rate after 6 months since the study ended at that point. Further, other potentially relevant factors for dropout were not available for analysis, including participant preferences, attitudes, and views about the app-based interventions, all of which may influence attrition [28]. These limitations should be considered in future research.

Conclusions

Dropout attrition was high in an RCT examining app-based interventions for unintentional child injury and parenting education versus only parenting education among caregivers of Chinese preschoolers. Engagement in the interventions differed across the two groups, and attrition was significantly associated with intervention engagement in both groups. Greater engagement with the app led to lower attrition. Attrition must be considered by researchers, policymakers, and practitioners when evaluating and implementing mHealth interventions. Efforts to engage users are critical to reduce attrition. Future research might consider feasible strategies to improve and maintain compliance of study participants to mHealth intervention programs.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (grant number: 81573260). The sponsors had no role in the design and conduct of the study; analysis, report, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. The authors thank Xiang Wei, Bo Chen, Jieyi He, Wangxin Xiao, Yuyan Gao, Liheng Tan, Yanhong Fu, Deyue Gao, and Cifu Xie from the Xiangya School of Public Health, Central South University, Changsha, China, for their assistance with the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Six-month attrition rates of caregivers of preschoolers (N=2920).
[DOCX File , 21 KB - jmir_v23i1e21636_app1.docx]

References

1. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
2. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
3. Silva BMC, Rodrigues JJPC, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
4. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845 [FREE Full text] [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
5. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of short message service text-based smoking cessation intervention among university students: a randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):321-328 [FREE Full text] [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
6. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
7. Liguori S, Stacchini M, Ciofi D, Olivini N, Bisogni S, Festini F. Effectiveness of an app for reducing preoperative anxiety in children: a randomized clinical trial. JAMA Pediatr 2016 Aug 01;170(8):e160533. [doi: 10.1001/jamapediatrics.2016.0533] [Medline: 27294708]
8. Ning P, Cheng P, Schwebel DC, Yang Y, Yu R, Deng J, et al. An app-based intervention for caregivers to prevent unintentional injury among preschoolers: cluster randomized controlled trial. JMIR Mhealth Uhealth 2019 Aug 09;7(8):e13519 [FREE Full text] [doi: 10.2196/13519] [Medline: 31400105]
9. Saposnik G, Cohen LG, Mamdani M, Pooyania S, Ploughman M, Cheung D, Stroke Outcomes Research Canada. Efficacy and safety of non-immersive virtual reality exercising in stroke rehabilitation (EVREST): a randomised, multicentre, single-blind, controlled trial. Lancet Neurol 2016 Sep;15(10):1019-1027 [FREE Full text] [doi: 10.1016/S1474-4422(16)30121-1] [Medline: 27365261]
10. Mirelman A, Rochester L, Maidan I, Del Din S, Alcock L, Nieuwhof F, et al. Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults (V-TIME): a randomised controlled trial. Lancet 2016 Sep 17;388(10050):1170-1182. [doi: 10.1016/S0140-6736(16)31325-3] [Medline: 27524393]
11. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893 [FREE Full text] [doi: 10.2196/10893] [Medline: 30355563]
12. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the ‘app’ in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014 Oct 29;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
13. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
14. Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol 2019 Dec 6;10:2745. [doi: 10.3389/fpsyg.2019.02745] [Medline: 31866915]
15. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
16. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
17. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

18. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
19. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014 Jul 02;9(1):11 [FREE Full text] [doi: 10.1186/1940-0640-9-11] [Medline: 24985342]
20. van Drongelen A, Boot CR, Hlobil H, Twisk JW, Smid T, van der Beek AJ. Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. Scand J Work Environ Health 2014 Nov;40(6):557-568 [FREE Full text] [doi: 10.5271/sjweh.3447] [Medline: 25121620]
21. Balk-Møller NC, Poulsen SK, Larsen TM. Effect of a nine-month web- and app-based workplace intervention to promote healthy lifestyle and weight loss for employees in the social welfare and health care sector: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e108 [FREE Full text] [doi: 10.2196/jmir.6196] [Medline: 28396303]
22. Burgess J, Watt K, Kimble RM, Cameron CM. Combining technology and research to prevent scald injuries (the cool runnings intervention): randomized controlled trial. J Med Internet Res 2018 Oct 10;20(10):e10361 [FREE Full text] [doi: 10.2196/10361] [Medline: 30305263]
23. Ning P, Chen B, Cheng P, Yang Y, Schwebel DC, Yu R, et al. Effectiveness of an app-based intervention for unintentional injury among caregivers of preschoolers: protocol for a cluster randomized controlled trial. BMC Public Health 2018 Jul 11;18(1):865 [FREE Full text] [doi: 10.1186/s12889-018-5790-1] [Medline: 29996813]
24. Changsha Municipal Bureau of Statistics, Changsha Investigation Team of the National Bureau of Statistics. In: Changsha Statistical Yearbook. Beijing: China Statistics Publishing House; Aug 2016.
25. Gielen AC, Bishai DM, Omaki E, Shields WC, McDonald EM, Rizzutti NC, et al. Results of an RCT in two pediatric emergency departments to evaluate the efficacy of an m-Health educational app on car seat use. Am J Prev Med 2018 Jun;54(6):746-755. [doi: 10.1016/j.amepre.2018.01.042] [Medline: 29656914]
26. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
27. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
28. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-S12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]

Abbreviations

aOR: adjusted odds ratio
mHealth: mobile health
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 19.06.20; peer-reviewed by P Akande, R Lystad; comments to author 15.07.20; revised version received 27.07.20; accepted 21.12.20; published 29.01.21.

Please cite as:
Li J, Ning P, Cheng P, Schwebel DC, Yang Y, Wei X, He J, Wang W, Li R, Hu G
Factors Associated With Dropout of Participants in an App-Based Child Injury Prevention Study: Secondary Data Analysis of a Cluster Randomized Controlled Trial
J Med Internet Res 2021;23(1):e21636
URL: https://www.jmir.org/2021/1/e21636
doi:10.2196/21636
PMID:33512318

Original Paper

Virtual Reality Relaxation for Patients With a Psychiatric Disorder: Crossover Randomized Controlled Trial

Wim Veling[1,2], MD, PhD; Bart Lestestuiver[1], BSc; Marieke Jongma[2], MSc; H J Rogier Hoenders[3], MD, PhD; Catheleine van Driel[1], MD, PhD

[1]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[2]VRelax BV, Groningen, Netherlands

[3]Centre for Integrative Psychiatry, Lentis, Groningen, Netherlands

Corresponding Author:
Wim Veling, MD, PhD
Department of Psychiatry
University Medical Center Groningen
University of Groningen
PO Box 30.001
Groningen, 9700 RB
Netherlands
Phone: 31 503612367
Email: w.veling@umcg.nl

Abstract

Background: Virtual reality (VR) relaxation is a promising mental health intervention that may be an effective tool for stress reduction but has hardly been tested in clinical trials with psychiatric patients. We developed an easy-to-use VR self-management relaxation tool (VRelax) with immersive 360° nature videos and interactive animated elements.

Objective: To investigate the immediate effects of VR relaxation on negative and positive affective states and short-term effects on perceived stress and symptoms in patients with a psychiatric disorder, compared to standard relaxation exercises.

Methods: A randomized crossover trial was conducted in 50 patients receiving ambulatory treatment for anxiety, psychotic, depressive, or bipolar disorder. Participants were randomly assigned to start with VRelax or standard relaxation and used both interventions for 10 days at home. They completed 8 visual analog scales of momentary negative and positive affective states before and after each session. Global perceived stress and psychiatric symptoms were measured before and after both intervention periods. Treatment effects were analyzed with multilevel mixed model regression analyses and 2-way analysis of variance.

Results: Both VRelax and standard relaxation exercises led to a statistically significant immediate improvement of all negative and positive affective states. Compared to standard relaxation, VRelax resulted in a significantly greater reduction of total negative affective state (change 16.2% versus 21.2%; $t_{1684}=-2.02$, 95% CI −18.70 to −0.28; $P=.04$). Specifically, VRelax had a stronger beneficial effect on momentary anxiety ($t_{1684}=-3.24$, 95% CI −6.86 to −1.69), sadness ($t_{1684}=-2.32$, 95% CI −6.51 to −0.55), and cheerfulness ($t_{1684}=2.35$, 95% CI 0.51 to 5.75). There were no significant differences between short-term effects of the two treatments on global perceived stress and symptoms.

Conclusions: If the results of this trial are replicated and extended, VRelax may provide a much-needed, effective, easy-to-use self-management relaxation intervention to enhance psychiatric treatments.

Trial Registration: Netherlands Trial Register NTR7294; https://www.trialregister.nl/trial/7096

(J Med Internet Res 2021;23(1):e17233) doi:10.2196/17233

KEYWORDS

virtual reality; stress; relaxation; negative affect; positive affect; depression; anxiety; randomized controlled trial

Introduction

Stress refers to the physiological, psychological, and behavioral responses to demands or perceived threats that challenge individuals' resources to manage them [1,2]. This study focuses on the psychological stress response, operationalized as the level of perceived distress and corresponding affective states (eg, feeling anxious, nervous, or down, and less calm, cheerful, or content).

Stress responses are often adaptive but also can increase vulnerability to disease. In psychiatry, stress is a transdiagnostic factor that has been related to both onset, course, and recurrence of mood, anxiety, and psychotic disorders [3,4]. Stress-reducing interventions such as breathing exercises, mental imagery, progressive muscle relaxation, and mindfulness are commonly applied in routine mental health care. These interventions are effective for reducing anxiety, stress, and sleeping problems [5,6], have some effects on level of depressive symptoms [7,8], and have hardly been investigated in patients with psychotic disorders [9]. Another intervention known to reduce stress is exposure to natural environments. Consistent associations have been reported between stress level (physiological and subjective) and time spent in green and blue spaces [10-12]. Negative impact of stress on health may be alleviated by presence of green space [13].

Although stress-reducing interventions may be effective, they are challenging for people with mental health problems. They require initiative, concentration, sustained attention, and energy, which are reduced in many psychiatric disorders, especially in association with exposure to stress [14]. There is a need for easy-to-use stress-reducing interventions that require less effort. Virtual reality (VR) technology offers opportunities to overcome the aforementioned challenges. A combination of visual and auditory stimuli in VR can be used to create an immersive experience that may reduce perceived distress and negative affect without much effort and induce relaxation and positive affect. VR treatments have been developed for various psychological and psychiatric problems, including anxiety and psychosis [15-17]. Most of these VR interventions are therapist-led and include exposure treatment, cognitive behavioral therapy, cognitive training, skills training, and stress management. Over the past decade, several studies have tested stress-reducing VR interventions in clinical and nonclinical samples [17]. A VR program for patients with a stress-related psychiatric disorder showed that confrontation with emotionally charged VR objects improved relaxation and negative mood significantly more than conventional cognitive behavioral therapy [18]. A VR intervention combining roleplays with teaching stress management techniques in a natural VR environment reduced perceived stress in workers with high-stress jobs [19]. A systematic review of VR stress management interventions in the military found 14 small studies, showing that VR stress inoculation decreased perceived stress and negative emotions [20]. Several studies specifically investigated the effect of VR relaxation on stress levels, both in healthy volunteers and people with psychiatric problems, mainly using exposure to VR natural environments [17]. Immersion in VR nature showed improvement of perceived stress, affective responses, and physiological stress measures, to a greater degree than neutral VR environments or two-dimensional nature videos [21-26]. These results suggest that VR-based stress management, in particular immersive VR nature relaxation, is a promising approach for stress reduction that merits further clinical research [17,27]. To the best of our knowledge, no randomized controlled trials investigated the effects of VR nature relaxation as a self-management intervention for stress reduction in patients with a psychiatric disorder. We developed VRelax, a VR self-management relaxation intervention with high-quality 360-degree videos of natural environments and interactive elements that enhance relaxation and focused attention to the environments (see Methods). In this study, we investigated the immediate effects of VRelax on negative and positive affective states of patients receiving outpatient treatment for a psychiatric disorder.

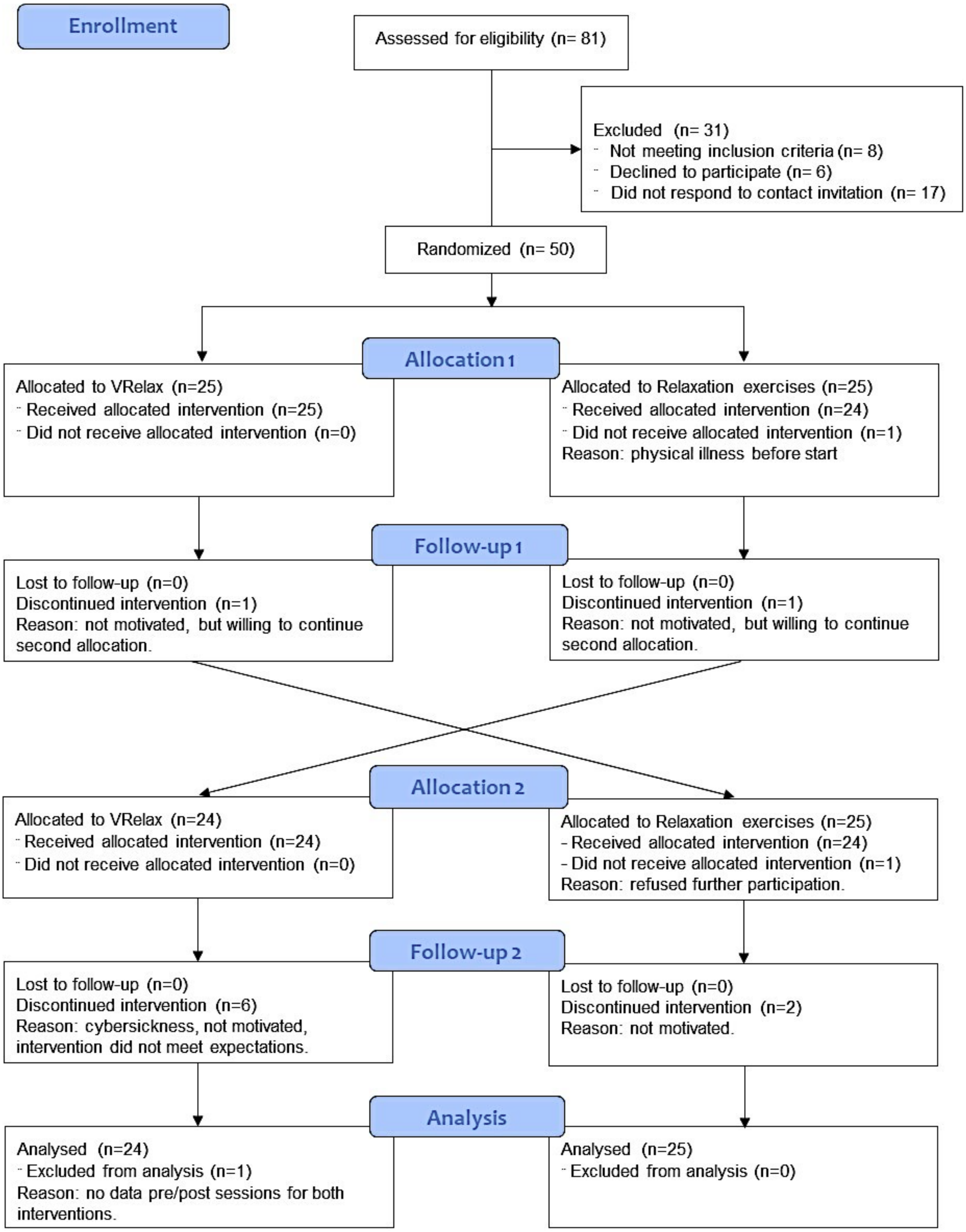
We conducted a crossover randomized controlled trial comparing VRelax with standard relaxation exercises. To evaluate immediate effects, momentary levels of negative and positive affective states were assessed before and after each session. As secondary outcomes, short-term effects (after 10 days) were measured with questionnaires on global perceived stress and psychiatric symptoms (anxiety, depression, paranoia). We hypothesized that (1a) VRelax results in an immediate improvement of negative and positive affective states, (1b) VRelax has a short-term effect on global perceived stress and psychiatric symptoms, and (2) the immediate and short-term effects of VRelax are stronger than the effects of standard relaxation exercises.

Methods

Participants

Patients receiving ambulatory treatment at University Medical Center Groningen (UMCG) Department of Psychiatry, Lentis Center for Integrative Psychiatry, or one of the two participating local general practices were eligible for the study. As stress is a major transdiagnostic factor in psychiatry, and psychiatric disorders often co-occur and partly overlap, a transdiagnostic approach was chosen [28]. Patients could participate if they had a diagnosis of depressive disorder, bipolar disorder, anxiety disorder, or psychotic disorder. Participants were referred to the study by their psychiatrist, psychologist, or general practitioner. A member of the study team contacted potential participants, provided information about the study, checked eligibility, and obtained informed consent. Patients reported their psychiatric diagnosis upon referral to the study; the clinicians who were involved in their regular treatment completed a brief checklist, including diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). The other inclusion criteria were perceived stress, based on self-report or clinician report; older than 18 years; and having a desktop computer or laptop computer at home. Exclusion criteria were a DSM-5 diagnosis of substance use disorder; benzodiazepine use greater than 10 mg per day of diazepam equivalent; diagnosis of epilepsy; and insufficient command of Dutch. A flowchart of the study participants is shown in Figure 1.

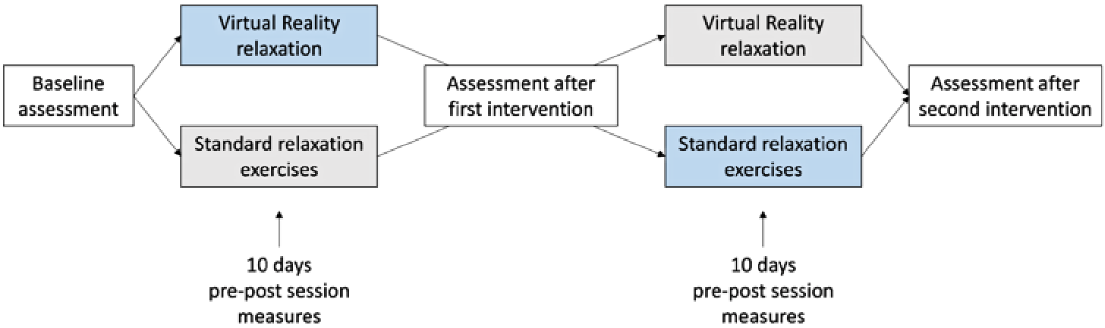
Figure 1. Flowchart of crossover randomized controlled trial.



Study Design and Procedures

The study was a randomized controlled crossover trial with two stress-reducing interventions: (1) the VRelax relaxation app and (2) standard relaxation exercises. Study procedures are shown in Figure 2. After informed consent, participants were randomly assigned to either start with VRelax or standard relaxation exercises. Randomization was conducted by an independent UMCG research coordinator using 1:1 randomization, drawing the assignment numbers not in series, but for each participant individually after informed consent was signed. After 10 days, they crossed over to the other intervention. Participants used both interventions consecutively for 10 days at home for a minimum of 10 minutes a day. Negative and positive affective states were assessed before and after each session; other measures were obtained at baseline (T0), after the first intervention period (T1), and after the second intervention period (T2). A crossover design was chosen because patients are very heterogeneous in clinical characteristics and behavior, and many confounding variables can be expected. As the primary outcome was immediate effect of interventions, carry-over effects were expected to be small. Ethical approval was obtained from the UMCG medical ethical committee (protocol number NL64380.042.17). The trial was registered prospectively in the Netherlands Trial Register (number NTR7294).

Figure 2. Study procedures.



Sample Size Calculation

Several studies provided preliminary evidence that VR relaxation reduces psychological and physiological stress, but to the best of our knowledge, no randomized controlled trials investigated the treatment effect of VR relaxation as a separate intervention in patients with a psychiatric disorder. As this was the first randomized VR stress reduction intervention study using VR nature relaxation and an active control condition, there was no estimation of effect size available. The sample size was therefore based on recommendations for a clinical pilot study and determined to be N=50 [29,30].

Interventions

All patients received regular treatment for their mental health problems throughout the study period. VRelax and standard relaxation were add-on interventions.

VRelax

The VRelax tool [31] was used with a Samsung Galaxy S6 or S7 smartphone, connected to a head-mounted display, the Samsung Gear VR. Three-dimensional audio was played with headphones. When VRelax was activated, the participant entered a white-walled waiting room. Presession measures (see Measures section) started automatically within the virtual waiting room. After completion of the presession measures, the walls of the waiting room disappeared, and the participant was standing on a beach, from which he or she could choose where to go. Participants could navigate between interconnected 360-degree video nature environments by looking at hotspots within their field of view that were activated after 2 seconds. VRelax included the following videos: several beaches, a coral reef with tropical fishes, swimming underwater with wild dolphins, a drone flight over a river landscape, a mountain meadow with animals, another mountain scenery in the Alps, a sea view from a cliff, and a beach session of Tibetan singing bowl therapy (Figure 3). Interactive elements were added as an extra layer on top of some videos; for example, audio tracks of guided meditation and progressive muscle relaxation could be activated by looking at hotspots; animated underwater floating air bubbles popped when looking at them; and a pattern of animated circles in the air produced a harmonious melody when looking at them in a particular order.

Figure 3. Impression of virtual relaxation environments from VRelax app.

The 360° videos were created by VIEMR, The Dolphin Swim Club, and Atmosphaeres; the VRelax tool was developed in collaboration with UMCG (first author WV).

Participants received a VRelax set to take home, including Gear VR, smartphone, and headphones. They were completely free to choose which video or videos within the app they preferred to use but were instructed to use the VRelax app at least once daily for 10 minutes. After 10 days, they returned the VR set to the study team.

Standard Relaxation

As active control intervention, participants received headphones and a USB stick with audio tracks of guided meditation and progressive relaxation exercises. They could play the tracks on devices they had at home (desktop or laptop computer, or smartphone). They were instructed to use it at least once daily for 10 minutes. After 10 days, participants returned the headphones and USB stick to the study team.

Measures

Primary Outcome

The primary outcome was the immediate effect on level of momentary negative and positive affective states, measured with visual analog scales (VAS; range: 0-100). The VAS items were scored immediately before and after each session, for VRelax within the virtual waiting room, and for the standard relaxation exercises with paper and pencil. Eight ecological momentary assessment items were selected that have been used in previous studies and have been related to psychological stress [32]. The following items were for negative affective states: I feel distressed; anxious; down; nervous (Cronbach α in our sample=.865). The following were for positive affective states: I feel relaxed; calm; cheerful; content (Cronbach α=.897).

Secondary Outcomes

Before the first intervention period, after the first intervention period, and after the second intervention period, patients completed questionnaires to assess short-term changes in perceived stress and psychiatric symptoms.

Perceived Stress

Global perceived level of stress over the past week was measured with the 10-item version of the Perceived Stress Scale [33], assessing the degree to which situations are appraised as stressful. The Perceived Stress Scale has good psychometric reliability and validity, including in samples of psychiatric patients [34]; in our sample, Cronbach α was .865.

Psychiatric Symptoms

Severity of depressive symptoms was measured with the Inventory of Depressive Symptomatology–Self-Report, a widely used 30-item questionnaire with good psychometric characteristics in psychiatric outpatients [35]. Cronbach α was .834 in the study sample. Level of anxiety symptoms was assessed with the 21-item Beck Anxiety Inventory, an instrument with high internal consistency, convergent and discriminant validity [36], and a Cronbach α of .908 in this study. With the Green Paranoid Thoughts Scale, ideas of social reference (16 items) and of persecution (16 items) were assessed [37]. This scale has been developed to assess paranoia over the spectrum of severity and has been recommended as the most reliable and valid self-report measure of paranoia [38]. In our sample, Cronbach α was .982.

Other Measures

Demographic and clinical information was obtained from participants: age, gender, marital status, level of education, work situation, DSM-5 classification, illness duration, and psychotropic medication use. During VRelax sessions, how long participants spent in each environment was automatically recorded in log files. Cybersickness was assessed to monitor adverse effects with the self-report Simulator Sickness Questionnaire (SSQ) before and after the VRelax intervention [39].

Statistical Analyses

All analyses were conducted in SPSS version 26 (IBM Corporation). Analyses of the primary outcome were done by the intention-to-treat approach.

Total negative affective state was calculated by summing the scores of the items anxious, down, distressed, and nervous. Similarly, the items relaxed, calm, cheerful, and content were summed for the total positive affective state. Changes of VAS items and total negative and positive affective state were calculated as 1 – (before session score / after session score). Descriptive statistics (mean, standard deviation, minimum and maximum scores) were calculated for all sociodemographic, clinical, and outcome variables.

As the primary outcome data had a multilevel structure (multiple VAS session scores within participants), VAS scores were analyzed with mixed model multilevel regression analyses (MIXED command), including the factors time (before vs after session), treatment (VRelax compared vs relaxation exercises), order of treatment administration (to adjust for carry-over effects), and the interaction time × treatment. Models had a random intercept for participant and a random slope for order of treatment administration. The estimation method was set to restricted maximum likelihood and the covariance structure to unstructured. First, the immediate effects of interventions were analyzed by comparing mean VAS scores before and after sessions for VRelax and relaxation exercises separately. Second, effects of VRelax compared to relaxation exercises were established by the time × treatment interaction.

Third, 2-way repeated-measures analysis of variance (ANOVA) was used to test differences between VRelax and standard relaxation exercises on global perceived stress, level of psychiatric symptoms, and cybersickness. Finally, carry-over effects were investigated by comparing outcome measures at baseline, T1, and T2 with paired *t* tests (baseline vs T1, T2 vs T1).

Results

Overview

Fifty patients were enrolled in this study between March 27 and December 17, 2018. One patient dropped out after randomization and did not start either condition, leaving 49 patients for analyses. For the primary outcome, 1.5% of data was missing (26 of 1762 VAS scores completed by 49 participants). Missing secondary outcome data ranged between 0% and 8.2%. Sociodemographic and clinical characteristics are shown in Table 1.

Table 1. Sociodemographic and clinical characteristics of the study sample.

Characteristic	Values
Participants, n	50
Age (years), mean (SD)	41.6 (14.2)
Sex, n (%)	
Male	17 (34)
Female	33 (66)
DSM-5 diagnosis, n (%)	
Bipolar disorder	13 (26)
Depressive disorder	16 (32)
Anxiety disorder	4 (8)
Psychotic disorder	6 (12)
Anxiety and depressive disorder	9 (18)
Anxiety and psychotic disorder	2 (4)
Duration psychiatric illness (years), mean (SD)	9.1 (9.4)
Marital status, n (%)	
Single	22 (44)
Relationship, not married	10 (20)
Married	15 (30)
Divorced	3 (6)
Level of education, n (%)	
Vocational	11 (22)
Secondary	2 (4)
Higher	37 (74)
Medication use, n (%)	
Antidepressant	20 (40)
Antipsychotic	13 (26)
Mood stabilizer	11 (22)
Benzodiazepine	17 (34)
Other psychotropic medication	2 (4)
No psychotropic medication	9 (18)

The mean number of sessions was 9.4 (SD 3.5; range: 0-16 sessions) for relaxation exercises and 8.7 (SD 4.4; range: 0-24 sessions) for VRelax; the difference was not statistically significant. VRelax log files showed that the average time spent in VRelax video environments was 17 minutes (SD 9.5) per session. Use of different VRelax environments is shown in Table 2. Environments with interactive elements and swimming with wild dolphins were used most frequently.

Table 2. Time spent in VRelax, by environment.

VRelax environment	Mean time spent, min:sec	Visits, n	Total time spent, min (%)
Beach with interactive relaxation exercises	8:51	143	1265 (27.7)
Scuba diving with dolphins	5:11	166	862 (18.9)
Sea view from cliff with interactive music element	3:32	136	480 (10.5)
Quiet beach, start position	1:44	275	477 (10.4)
Coral reef with interactive air bubble game	2:39	150	398 (8.7)
Session of Tibetan sound therapy	3:39	88	321 (7.0)
Quiet beach with rocks	1:52	127	238 (5.2)
Mountain scenery without animals	2:17	86	195 (4.3)
Drone flight over river landscape	2:29	67	167 (3.7)
Mountain meadow with cows	2:23	68	162 (3.5)

Primary Outcome: Immediate Effects

Both VRelax and standard relaxation exercises led to a statistically significant reduction of negative affective states and improvement of positive affective states, as indicated by improvements on all VAS items (Tables 3 and 4). Mean changes in VAS item scores from before to after sessions within individuals were between 11.4%and 28.2% with VRelax, and between 4.6% and 28.3% with relaxation exercises (Table 5). VRelax reduced negative affective states significantly more than relaxation exercises (16.2% versus 21.2%; $t_{1684}=-2.02$, 95% CI −18.70 to −0.28; *P*=.04). Of the separate momentary affective state items, VRelax had a larger beneficial effect than relaxation exercises on feeling anxious, down, or cheerful.

Table 3. Effects of VRelax on momentary psychological stress.[a]

VAS[b] items	Before, mean (SD)	After, mean (SD)	Difference		
			t	95% CI	*P* value
Relaxed	44.0 (21.2)	56.4 (22.8)	11.71	10.46 to 14.67	<.001
Calm	44.8 (20.9)	56.1 (23.6)	10.73	9.44 to 13.67	<.001
Cheerful	40.3 (24.9)	44.9 (25.4)	4.92	2.94 to 6.86	<.001
Content	42.2 (23.3)	49.8 (24.8)	8.41	6.00 to 9.65	<.001
Total positive affective state	171.3 (79.0)	207.1 (84.9)	11.11	30.35 to 43.37	<.001
Distressed	45.5 (23.7)	34.7 (22.8)	−9.42	−12.97 to −8.49	<.001
Anxious	31.4 (23.9)	23.8 (22.4)	−7.57	−9.57 to −5.63	<.001
Down	43.5 (24.2)	35.7 (23.2)	−7.10	−10.14 to −5.75	<.001
Nervous	39.2 (24.3)	31.6 (23.1)	−7.32	−9.56 to −5.52	<.001
Total negative affective state	159.6 (79.9)	125.8 (78.2)	−9.65	−40.69 to −26.93	<.001

[a] Mixed model regression analyses, comparison of mean scores after and before sessions, adjusted for order of interventions.

[b] VAS: visual analog scales.

Table 4. Effects of relaxation exercises on momentary psychological stress.[a]

VAS[b] items	Before, mean (SD)	After, mean (SD)	Difference		
			t	95% CI	*P* value
Relaxed	44.8 (20.0)	56.8 (19.6)	12.15	10.01 to 13.87	<.001
Calm	42.8 (20.0)	54.9 (19.7)	11.47	9.96 to 14.07	<.001
Cheerful	39.5 (21.8)	41.3 (21.8)	2.19	0.18 to 3.35	.03
Content	41.9 (20.7)	48.2 (20.3)	7.24	4.48 to 7.81	<.001
Total positive affective state	169.0 (72.2)	200.9 (70.2)	10.74	25.94 to 37.54	<.001
Distressed	44.4 (23.5)	34.5 (23.1)	-9.05	−12.13 to −7.80	<.001
Anxious	27.8 (25.3)	24.4 (23.9)	-4.68	−4.81 to −1.97	<.001
Down	40.9 (25.1)	36.6 (24.8)	-4.38	−6.11 to −2.65	<.001
Nervous	36.6 (25.1)	30.0 (23.1)	-7.24	−8.38 to −4.80	<.001
Total negative affective state	149.7 (81.8)	125.4 (81.2)	-9.10	−29.62 to −19.11	<.001

[a] Mixed model regression analyses, comparison of mean scores after and before sessions, adjusted for order of interventions.

[b] VAS: visual analog scales.

Table 5. Comparison of treatment effects between VRelax and relaxation exercises.

VAS[a] items	VRelax change, %[b]	Relaxation exercises change, %	Test differences between interventions[c]			
			F	*t*	95% CI	*P* value
Relaxed	28.2	26.8	0.142	0.38	−2.40 to 3.54	.71
Calm	25.2	28.3	0.104	-0.32	−3.52 to 2.53	.75
Cheerful	11.4	4.6	5.510	2.35	0.51 to 5.75	.02
Content	18.0	15.0	0.422	1.24	−0.96 to 4.28	.22
Total positive affective state	20.9	18.9	1.16	1.08	−4.14 to 14.19	.28
Distressed	−23.7	−22.3	0.221	−0.47	−4.04 to 2.48	.64
Anxious	−24.2	−12.2	10.53	−3.24	−6.86 to −1.69	.001
Down	−17.9	−10.5	5.39	−2.32	−6.51 to −0.55	.02
Nervous	−19.4	−18.0	1.537	−0.65	−3.77 to 1.90	.52
Total negative affective state	−21.2	−16.2	4.082	−2.02	−18.70 to −0.28	.04

[a] VAS: visual analog scales.

[b] Percentage of change, mean scores after sessions compared to scores before sessions.

[c] Mixed model regression analyses. Treatment effect is estimated as interaction between time (before/after sessions) and type of intervention, adjusted for order of interventions.

Secondary Outcomes: Short-term Effects

Short-term effects on perceived stress and symptoms are shown in Table 6. Symptoms of depression and anxiety were reduced significantly after VRelax use. Perceived global stress level, symptoms of depression, and paranoid thoughts were significantly lower after standard relaxation exercises. There were no significant differences between effects of the two treatments, only a trend toward a superior effect of VRelax on anxiety symptoms (interaction time × treatment, F_1=3.650; *P*=.06).

We found period effects on momentary mood states and psychiatric symptoms (Table 7), with stronger effects in the second period than in the first.

No serious adverse events occurred. Several participants reported cybersickness, and 2 stopped using VRelax because of nausea and dizziness. Mean total score on the SSQ, however, was lower rather than higher after VRelax (43.1, SD 10.9) compared to before VRelax (48.3, SD 12.7).

Table 6. Perceived stress, psychiatric symptoms, and cybersickness before and after treatments.[a]

Measure	VRelax				Standard relaxation exercises				Difference between treatments[b]	
	Before, mean (SD)	After, mean (SD)	Paired *t* test	*P* value	Before, mean (SD)	After, mean (SD)	Paired *t* test	*P* value	*F* test (*df*)	*P* value
PSS[c]	31.4 (7.5)	30.1 (6.7)	1.65	.11	31.8 (7.3)	29.5 (7.1)	2.77	.01	0.527 (1)	.47
IDS-SR[d]	39.3 (13.6)	35.8 (13.0)	3.31	.002	38.9 (14.4)	36.4 (14.3)	2.34	.02	0.519 (1)	.48
BAI[e]	20.7 (12.3)	16.0 (10.2)	4.30	<.001	18.5 (11.4)	16.8 (12.8)	1.68	.10	3.650 (1)	.06
GPTS[f]	45.8 (21.7)	46.4 (28.1)	0.03	.98	50.1 (29.7)	44.4 (24.1)	2.07	.04	0.760 (1)	.39

[a] Samples sizes of comparisons range between 49 and 45, due to incomplete data.

[b] Two-way ANOVA, interaction term time × treatment.

[c] PSS: Perceived Stress Scale.

[d] IDS-SR: Inventory of Depressive Symptomatology–Self-Report.

[e] BAI: Beck Anxiety Inventory.

[f] GPTS: Green Paranoid Thoughts Scale.

Table 7. Changes in momentary affective states, perceived stress and psychiatric symptoms over time.[a]

Measure	T0 (baseline)	T1		T2	
	Mean (SD)	Mean (SD)	*P* value	Mean (SD)	*P* value
Change negative momentary affective states[b]	N/A[c]	−24.7 (43.8)	N/A	−33.2 (38.1)	.003
Change positive momentary affective states[b]	N/A	29.8 (46.7)	N/A	39.6 (40.3)	.001
PSS[d,e]	32.7 (7.5)	30.4 (7.3)	.001	28.8 (6.5)	.22
IDS-SR[e,f]	41.1 (13.7)	36.3 (13.4)	<.001	34.5 (12.5)	.35
BAI[e,g]	21.3 (12.0)	18.0 (11.7)	.002	14.7 (11.2)	.01
GPTS[e,h]	49.8 (27.9)	43.9 (21.0)	.03	45.2 (28.4)	.87

[a] Sample sizes of comparisons range between 49 and 45, due to incomplete data.

[b] Change in mean before/after session scores. T1 and T2 are different subjects because of crossover study design; *P* values for independent *t* tests.

[c] N/A: not applicable.

[d] PSS: Perceived Stress Scale.

[e] Repeated measures; *P* values for paired *t* test, T1 compared to baseline or T2 compared to T1.

[f] IDS-SR: Inventory of Depressive Symptomatology–Self-Report.

[g] BAI: Beck Anxiety Inventory.

[h] GPTS: Green Paranoid Thoughts Scale.

Discussion

This randomized crossover clinical trial found that VR relaxation immediately reduced negative affective states and improved positive affective states in patients who concurrently received ambulatory psychiatric treatment. VRelax had a stronger effect on negative affective states than standard relaxation exercises, in particular on feeling anxious, down, or cheerful. Psychiatric symptoms, measured over a 10-day period, improved somewhat in both conditions.

The results suggest that immersive VR nature exposure is an effective tool for immediate improvement of affective states and reduction of psychological stress, more powerful than conventional relaxation exercises for making people feel better. The standard relaxation exercises were audio tracks that participants played on their computer or smartphone. Some tracks included 2-dimensional nature pictures. Numerous VR studies have shown that immersion in a virtual environment and the experience of being in a different world strongly induce psychological and physical effects [40]. Although viewing 2-dimensional nature images and videos has previously also been related to stress reduction [41], the immersive nature experience of VR is likely to have contributed to the superior effect.

The immersive VR experience is likely to distract attention from negative thoughts and stimuli, as is found in VR pain distraction studies [42,43]. Attention restoration has been proposed as a key mechanism of health-promoting effects of nature exposure [44-46]. Attention fatigue occurs when, after prolonged and intense use, the capacity to direct voluntary attention is reduced, and the capacity to ward off negative internal or external distractions is reduced [47]. The salient and pleasant characteristics of natural environments capture attention automatically and effortlessly, enabling recovery of voluntary, cognitively directed attention [11,48]. Another potential mechanism originates from psycho-evolutionary theory, stating that exposure to natural environments rapidly restores positive affect and reduces stress-related physiological activation because in earlier times, rapid recovery of stress increased chances of survival [49]. In both theories, nature exposure leads to positive affect and stress reduction, in accordance with the results that were found in the present study.

The effect of VR relaxation was not so strong on global perceived stress over the past 10 days. The perceived stress measure in this study included other aspects of psychological stress besides negative affective states, such as level of perceived control over things that happen, confidence to cope with problems, and ability to overcome difficulties [33]. Whereas affective states are sensitive to immediate and short-term changes [32], the cognitive and behavioral components of stress are likely to change more gradually. Optimal duration of VR relaxation for this purpose may be longer than 10 days, or the study period may have been too short to capture such changes.

Capturing and keeping attention is likely to be more difficult in patients with psychiatric disorders than in healthy individuals because vigilance is impaired and their cognitively directed attention is more severely fatigued [50]. Feedback of psychiatric inpatients during the prototyping phase of VRelax suggested that repeated passive viewing of calm, uneventful virtual nature environments was not always sufficient to distract from negative thoughts and feelings. We therefore added interactive elements to some of the videos, (ie, a relaxation exercise, a simple game of air bubble popping, and musical tones played in response to head movements). Indeed, in this study, the VRelax log files showed that the interactive videos were used longer and more frequently than the “passive” videos.

There was one notable exception: scuba diving with wild dolphins had the highest number of visits. The popularity of the dolphin video can be explained by several factors. Throughout the 7-minute video, there is a lot to be seen all around, with more than 20 active dolphins constantly moving, which may have created a stronger sense of immersion than the other videos. In addition, health-promoting effects on humans have been ascribed to dolphins. Dolphin-assisted therapy is offered for a range of physical and psychological problems, including autism and depression, albeit scientific evidence for efficacy is limited and debated [51,52].

Participants did not report many problems with the use of VR. Only 2 patients stopped using VRelax because of cybersickness, and SSQ scores were low after the intervention. It has been suggested that patients with a psychiatric disorder may experience increased psychiatric symptoms when using VR [16,53]. This may be caused by challenging VR content such as exposure to anxiety-evoking stimuli, which was not part of VRelax. Another potential cause of anxiety and paranoia may be that wearing a head-mounted display disconnects people from their real environment. These issues did not occur during the study, perhaps because participants used VRelax in the safe environment of their own home.

Strengths of this study include the randomized crossover design. Participants served as their own control, minimizing the impact of confounding factors. Analyses of temporal effects suggest carry-over effects of both interventions. Effects on primary and secondary outcomes were stronger in the second intervention period than in the first. However, the impacts of carry-over or period effects on the results of the study are likely to be small because participants were randomized in a 1:1 ratio for order of interventions, and analyses of the primary outcome were adjusted for order of interventions. As a consequence of the crossover design, it was not possible to investigate the effects of VR relaxation on medium- or long-term outcomes. It is conceivable that VR experiences have a longer-lasting impact than standard relaxation exercises because of the immersive character, or simply because people may find it easier to keep using VRelax than to keep doing relaxation exercises. We did not observe such a pattern during the short duration of this study.

Patients used the interventions multiple times over a (double) 10-day period, with effect measures before and after each session. These frequent assessments and detailed longitudinal data increased real life ecological validity and reliability of the primary outcome measure and increased statistical power of the study. A disadvantage was that patients had to reflect on their affective state immediately after relaxation sessions, which may have reduced the beneficial effects to some extent.

Another limitation is that we did not measure physiological effects of VR relaxation, which would have been a valuable outcome measure and a possible way to study biological mechanisms of VR nature exposure. Relaxation therapy, as well as nature exposure, has been related to changes in physiological stress parameters, such as vagal tone, electroencephalography alpha and theta waves, heart rate variability, blood pressure, gene expression, and cortisol level [54,55]. Recent research of VR nature videos found that electroencephalography event-related potentials may be used as a marker of attention restoration [56].

If the results of this trial are replicated and extended, VR nature exposure may provide an effective relaxation intervention to support and enhance psychiatric treatments. Stress management is a cornerstone of psychiatric therapies, but there is a dearth of effortless, pleasant, and powerful self-management tools for improvement of affective states and reduction of stress. Longer-term effects of VR relaxation on perceived stress, psychiatric symptoms, and relapse prevention should be investigated, as well as cost-effectiveness. In addition, integration of biofeedback and more interactive elements (eg, breathing exercises) may increase efficacy of VRelax. Finally, a broader use of VR relaxation can be envisioned for people in need of distraction, for example, during unpleasant medical interventions, during long isolated hospital stays, or for bedridden (terminally) ill patients. VR applications are rapidly growing and may have a profound impact on the landscape of mental health interventions.

Acknowledgments

Funding for this study was provided by UMCG, Healthy Ageing grant, number CDO17.0014/2017-1/308. The authors wish to thank The Dolphin Swim Club for sharing the 360-degree dolphin video and for their contribution to the development of the VRelax tool.

All statistical analyses were independently checked by Dr Klaas Wardenaar, statistician at University of Groningen, UMCG.

Authors' Contributions

WV designed the study, obtained funding, analyzed the data, and drafted the manuscript. BL contributed to the development of the interventions, collected the data, and critically reviewed the manuscript. MJ contributed to development of the VRelax tool, data collection, and analyses and commented on the manuscript. RH and CvD contributed to data collection, interpretation of the data, and writing and editing of the manuscript.

Conflicts of Interest

WV is cofounder and chief scientific officer of VRelax BV, the company that has developed the VR in collaboration with UMCG, and holds shares in VRelax BV. MJ is cofounder of VRelax BV and holds shares. The remaining authors have no conflicts to declare.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1764 KB - jmir_v23i1e17233_app1.pdf]

References

1. Selye H. The stress of life. New York: McGraw-Hill; 1956.
2. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
3. Monroe SM, Harkness KL. Life stress, the "kindling" hypothesis, and the recurrence of depression: considerations from a life stress perspective. Psychol Rev 2005 Apr;112(2):417-445. [doi: 10.1037/0033-295X.112.2.417] [Medline: 15783292]
4. Phillips LJ, Francey SM, Edwards J, McMurray N. Stress and psychosis: towards the development of new models of investigation. Clin Psychol Rev 2007 Apr;27(3):307-317. [doi: 10.1016/j.cpr.2006.10.003] [Medline: 17169470]
5. Manzoni GM, Pagnini F, Castelnuovo G, Molinari E. Relaxation training for anxiety: a ten-years systematic review with meta-analysis. BMC Psychiatry 2008;8:41 [FREE Full text] [doi: 10.1186/1471-244X-8-41] [Medline: 18518981]
6. Strauss C, Cavanagh K, Oliver A, Pettman D. Mindfulness-based interventions for people diagnosed with a current episode of an anxiety or depressive disorder: a meta-analysis of randomised controlled trials. PLoS One 2014;9(4):e96110 [FREE Full text] [doi: 10.1371/journal.pone.0096110] [Medline: 24763812]
7. Jorm AF, Morgan AJ, Hetrick SE. Relaxation for depression. Cochrane Database Syst Rev 2008(4):CD007142. [doi: 10.1002/14651858.CD007142.pub2] [Medline: 18843744]
8. Kuyken W, Warren FC, Taylor RS, Whalley B, Crane C, Bondolfi G, et al. Efficacy of Mindfulness-Based Cognitive Therapy in Prevention of Depressive Relapse: An Individual Patient Data Meta-analysis From Randomized Trials. JAMA Psychiatry 2016 Jun 01;73(6):565-574. [doi: 10.1001/jamapsychiatry.2016.0076] [Medline: 27119968]
9. Vancampfort D, Correll CU, Scheewe TW, Probst M, De Herdt A, Knapen J, et al. Progressive muscle relaxation in persons with schizophrenia: a systematic review of randomized controlled trials. Clin Rehabil 2013 Apr;27(4):291-298. [doi: 10.1177/0269215512455531] [Medline: 22843353]
10. de Vries S, Ten Have M, van Dorsselaer S, van Wezep M, Hermans T, de Graaf R. Local availability of green and blue space and prevalence of common mental disorders in the Netherlands. BJPsych Open 2016 Nov;2(6):366-372 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002469] [Medline: 29018562]
11. Hartig T, Mitchell R, de Vries S, Frumkin H. Nature and health. Annu Rev Public Health 2014;35:207-228. [doi: 10.1146/annurev-publhealth-032013-182443] [Medline: 24387090]
12. Song C, Ikei H, Miyazaki Y. Physiological Effects of Nature Therapy: A Review of the Research in Japan. Int J Environ Res Public Health 2016 Aug 03;13(8):781 [FREE Full text] [doi: 10.3390/ijerph13080781] [Medline: 27527193]
13. Van den Berg AE. From Green Space to Green Prescriptions: Challenges and Opportunities for Research and Practice. Front Psychol 2017;8:268 [FREE Full text] [doi: 10.3389/fpsyg.2017.00268] [Medline: 28289398]
14. Millan MJ, Agid Y, Brüne M, Bullmore ET, Carter CS, Clayton NS, et al. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. Nat Rev Drug Discov 2012 Feb;11(2):141-168. [doi: 10.1038/nrd3628] [Medline: 22293568]

15. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400. [doi: 10.1017/S003329171700040X] [Medline: 28325167]
16. Park MJ, Kim DJ, Lee U, Na EJ, Jeon HJ. A Literature Overview of Virtual Reality (VR) in Treatment of Psychiatric Disorders: Recent Advances and Limitations. Front Psychiatry 2019;10:505 [FREE Full text] [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
17. Pizzoli SFM, Mazzocco K, Triberti S, Monzani D, Alcañiz Raya ML, Pravettoni G. User-Centered Virtual Reality for Promoting Relaxation: An Innovative Approach. Front Psychol 2019;10:479 [FREE Full text] [doi: 10.3389/fpsyg.2019.00479] [Medline: 30914996]
18. Baños R, Guillen V, Quero S, García-Palacios A, Alcaniz M, Botella C. A virtual reality system for the treatment of stress-related disorders: A preliminary analysis of efficacy compared to a standard cognitive behavioral program. International Journal of Human-Computer Studies 2011 Aug;69(9):602-613. [doi: 10.1016/j.ijhcs.2011.06.002]
19. Gaggioli A, Pallavicini F, Morganti L, Serino S, Scaratti C, Briguglio M, et al. Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. J Med Internet Res 2014;16(7):e167 [FREE Full text] [doi: 10.2196/jmir.3235] [Medline: 25004803]
20. Pallavicini F, Argenton L, Toniazzi N, Aceti L, Mantovani F. Virtual Reality Applications for Stress Management Training in the Military. Aerosp Med Hum Perform 2016 Dec 01;87(12):1021-1030. [doi: 10.3357/AMHP.4596.2016] [Medline: 28323588]
21. Valtchanov D, Ellard C. Physiological and affective responses to immersion in virtual reality: effects of nature and urban settings. J Cyber Ther Rehabil 2010;3:359-373.
22. Villani D, Riva G. Does interactive media enhance the management of stress? Suggestions from a controlled study. Cyberpsychol Behav Soc Netw 2012 Jan;15(1):24-30. [doi: 10.1089/cyber.2011.0141] [Medline: 22032797]
23. Valtchanov D, Barton KR, Ellard C. Restorative effects of virtual nature settings. Cyberpsychol Behav Soc Netw 2010 Oct;13(5):503-512. [doi: 10.1089/cyber.2009.0308] [Medline: 20950174]
24. Anderson AP, Mayer MD, Fellows AM, Cowan DR, Hegel MT, Buckey JC. Relaxation with Immersive Natural Scenes Presented Using Virtual Reality. Aerosp Med Hum Perform 2017 Jun 01;88(6):520-526. [doi: 10.3357/AMHP.4747.2017] [Medline: 28539139]
25. Liszio S, Graf L, Masuch M. The relaxing effect of virtual nature: Immersive technology provides relief in acute stress situations. Annu Rev CyberTherapy Telemed 2018;16:87-93.
26. Shah LBI, Torres S, Kannusamy P, Chng CML, He H, Klainin-Yobas P. Efficacy of the virtual reality-based stress management program on stress-related variables in people with mood disorders: the feasibility study. Arch Psychiatr Nurs 2015 Feb;29(1):6-13. [doi: 10.1016/j.apnu.2014.09.003] [Medline: 25634868]
27. Annerstedt M, Jönsson P, Wallergård M, Johansson G, Karlson B, Grahn P, et al. Inducing physiological stress recovery with sounds of nature in a virtual reality forest--results from a pilot study. Physiol Behav 2013 Jun 13;118:240-250. [doi: 10.1016/j.physbeh.2013.05.023] [Medline: 23688947]
28. Fusar-Poli P, Solmi M, Brondino N, Davies C, Chae C, Politi P, et al. Transdiagnostic psychiatry: a systematic review. World Psychiatry 2019 Jun;18(2):192-207 [FREE Full text] [doi: 10.1002/wps.20631] [Medline: 31059629]
29. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiol 2012 Mar;65(3):301-308. [doi: 10.1016/j.jclinepi.2011.07.011] [Medline: 22169081]
30. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
31. VRelax. URL: https://www.vrelax.com [accessed 2020-12-30]
32. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van OJ. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med 2009 Sep;39(9):1533-1547. [doi: 10.1017/S0033291708004947] [Medline: 19215626]
33. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
34. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
35. Rush AJ, Carmody T, Reimitz P. The Inventory of Depressive Symptomatology (IDS): Clinician (IDS-C) and Self-Report (IDS-SR) ratings of depressive symptoms. Int J Method Psychiat Res 2006 Jun;9(2):45-59. [doi: 10.1002/mpr.79]
36. Fydrich T, Dowdall D, Chambless DL. Reliability and validity of the beck anxiety inventory. Journal of Anxiety Disorders 1992 Jan;6(1):55-61. [doi: 10.1016/0887-6185(92)90026-4]
37. Green CEL, Freeman D, Kuipers E, Bebbington P, Fowler D, Dunn G, et al. Measuring ideas of persecution and social reference: the Green et al. Paranoid Thought Scales (GPTS). Psychol Med 2008 Jan;38(1):101-111. [doi: 10.1017/S0033291707001638] [Medline: 17903336]
38. Statham V, Emerson L, Rowse G. A systematic review of self-report measures of paranoia. Psychol Assess 2019 Mar;31(2):139-158. [doi: 10.1037/pas0000645] [Medline: 30234319]

39. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. The International Journal of Aviation Psychology 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
40. Sanchez-Vives MV, Slater M. From presence to consciousness through virtual reality. Nat Rev Neurosci 2005 Apr;6(4):332-339. [doi: 10.1038/nrn1651] [Medline: 15803164]
41. van den Berg MMHE, Maas J, Muller R, Braun A, Kaandorp W, van Lien R, et al. Autonomic Nervous System Responses to Viewing Green and Built Settings: Differentiating Between Sympathetic and Parasympathetic Activity. Int J Environ Res Public Health 2015 Dec 14;12(12):15860-15874 [FREE Full text] [doi: 10.3390/ijerph121215026] [Medline: 26694426]
42. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
43. Tashjian VC, Mosadeghi S, Howard AR, Lopez M, Dupuy T, Reid M, et al. Virtual Reality for Management of Pain in Hospitalized Patients: Results of a Controlled Trial. JMIR Ment Health 2017 Mar 29;4(1):e9 [FREE Full text] [doi: 10.2196/mental.7387] [Medline: 28356241]
44. Kaplan S. The restorative benefits of nature: Toward an integrative framework. Journal of Environmental Psychology 1995 Sep;15(3):169-182. [doi: 10.1016/0272-4944(95)90001-2]
45. Berman MG, Jonides J, Kaplan S. The cognitive benefits of interacting with nature. Psychol Sci 2008 Dec;19(12):1207-1212. [doi: 10.1111/j.1467-9280.2008.02225.x] [Medline: 19121124]
46. Pasanen T, Johnson K, Lee K, Korpela K. Can Nature Walks With Psychological Tasks Improve Mood, Self-Reported Restoration, and Sustained Attention? Results From Two Experimental Field Studies. Front Psychol 2018;9:2057 [FREE Full text] [doi: 10.3389/fpsyg.2018.02057] [Medline: 30425671]
47. Berto R. Exposure to restorative environments helps restore attentional capacity. Journal of Environmental Psychology 2005 Sep;25(3):249-259. [doi: 10.1016/j.jenvp.2005.07.001]
48. Sahlin E, Lindegård A, Hadzibajramovic E, Grahn P, Vega Matuszczyk J, Ahlborg G. The Influence of the Environment on Directed Attention, Blood Pressure and Heart Rate—An Experimental Study Using a Relaxation Intervention. Landscape Research 2014 Dec 22;41(1):7-25. [doi: 10.1080/01426397.2014.982079]
49. Ulrich RS, Simons RF, Losito BD, Fiorito E, Miles MA, Zelson M. Stress recovery during exposure to natural and urban environments. Journal of Environmental Psychology 1991 Sep;11(3):201-230. [doi: 10.1016/s0272-4944(05)80184-7]
50. Bowie CR, Milanovic M, Tran T, Cassidy S. Disengagement from tasks as a function of cognitive load and depressive symptom severity. Cogn Neuropsychiatry 2017 Jan;22(1):83-94. [doi: 10.1080/13546805.2016.1267617] [Medline: 27996635]
51. Antonioli C, Reveley MA. Randomised controlled trial of animal facilitated therapy with dolphins in the treatment of depression. BMJ 2005 Nov 26;331(7527):1231 [FREE Full text] [doi: 10.1136/bmj.331.7527.1231] [Medline: 16308382]
52. Fiksdal BL, Houlihan D, Barnes AC. Dolphin-Assisted Therapy: Claims versus Evidence. Autism Res Treat 2012;2012:839792 [FREE Full text] [doi: 10.1155/2012/839792] [Medline: 22928101]
53. Rus-Calafell M, Garety P, Sason E, Craig TJK, Valmaggia LR. Virtual reality in the assessment and treatment of psychosis: a systematic review of its utility, acceptability and effectiveness. Psychol Med 2018 Feb;48(3):362-391. [doi: 10.1017/S0033291717001945] [Medline: 28735593]
54. Bhasin MK, Dusek JA, Chang B, Joseph MG, Denninger JW, Fricchione GL, et al. Relaxation response induces temporal transcriptome changes in energy metabolism, insulin secretion and inflammatory pathways. PLoS One 2013;8(5):e62817 [FREE Full text] [doi: 10.1371/journal.pone.0062817] [Medline: 23650531]
55. Furuyashiki T, Kitaoka S. Neural mechanisms underlying adaptive and maladaptive consequences of stress: Roles of dopaminergic and inflammatory responses. Psychiatry Clin Neurosci 2019 Nov;73(11):669-675. [doi: 10.1111/pcn.12901] [Medline: 31215710]
56. Chung K, Lee D, Park JY. Involuntary Attention Restoration During Exposure to Mobile-Based 360° Virtual Nature in Healthy Adults With Different Levels of Restorative Experience: Event-Related Potential Study. J Med Internet Res 2018 Nov 30;20(11):e11152 [FREE Full text] [doi: 10.2196/11152] [Medline: 30504121]

Abbreviations

ANOVA: analysis of variance
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
SSQ: Simulator Sickness Questionnaire
UMCG: University Medical Center Groningen
VAS: visual analog scales
VR: virtual reality

Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by F Pallavicini, P Lindner; comments to author 10.02.20; revised version received 29.05.20; accepted 25.08.20; published 15.01.21.

Please cite as:
Veling W, Lestestuiver B, Jongma M, Hoenders HJR, van Driel C
Virtual Reality Relaxation for Patients With a Psychiatric Disorder: Crossover Randomized Controlled Trial
J Med Internet Res 2021;23(1):e17233
URL: http://www.jmir.org/2021/1/e17233/
doi:10.2196/17233
PMID:33448933

Original Paper

Virtual Reality Cognitive Training Among Individuals With Alcohol Use Disorder Undergoing Residential Treatment: Pilot Randomized Controlled Trial

Pedro Gamito[1,2], DPhil; Jorge Oliveira[1,2], DPhil; Marcelo Matias[1], MSc; Elsa Cunha[1], MSc; Rodrigo Brito[2], DPhil; Paulo Ferreira Lopes[1,2], DPhil; Alberto Deus[3], DPhil

[1]School of Psychology and Life Sciences, Lusófona University, Lisboa, Portugal

[2]Digital Human-Environment Interaction Lab, Lusófona University, Lisboa, Portugal

[3]Casa de Saúde do Telhal, Instituto São João de Deus, Sintra, Portugal

Corresponding Author:
Jorge Oliveira, DPhil
School of Psychology and Life Sciences
Lusófona University
Campo Grande, 376
Lisboa, 1749-024
Portugal
Phone: 351 217515500
Email: jorge.oliveira@ulusofona.pt

Abstract

Background: Alcohol use disorder (AUD) has been associated with diverse physical and mental morbidities. Among the main consequences of chronic and excessive alcohol use are cognitive and executive deficits. Some of these deficits may be reversed in specific cognitive and executive domains with behavioral approaches consisting of cognitive training. The advent of computer-based interventions may leverage these improvements, but randomized controlled trials (RCTs) of digital interactive-based interventions are still scarce.

Objective: The aim of this study is to explore whether a cognitive training approach using VR exercises based on activities of daily living is feasible for improving the cognitive function of patients with AUD undergoing residential treatment, as well as to estimate the effect size for this intervention to power future definitive RCTs.

Methods: This study consisted of a two-arm pilot RCT with a sample of 36 individuals recovering from AUD in a therapeutic community; experimental group participants received a therapist-guided, VR-based cognitive training intervention combined with treatment as usual, and control group participants received treatment as usual without cognitive training. A comprehensive neuropsychological battery of tests was used both at pre- and postassessments, including measurement of global cognition, executive functions, attention, visual memory, and cognitive flexibility.

Results: In order to control for potential effects of global cognition and executive functions at baseline, these domains were controlled for in the statistical analysis for each individual outcome. Results indicate intervention effects on attention in two out of five outcomes and on cognitive flexibility in two out of six outcomes, with effect sizes in significant comparisons being larger for attention than for cognitive flexibility. Patient retention in cognitive training was high, in line with previous studies.

Conclusions: Overall, the data suggest that VR-based cognitive training results in specific contributions to improving attention ability and cognitive flexibility of patients recovering from AUD.

Trial Registration: ClinicalTrials.gov NCT04505345; https://clinicaltrials.gov/show/NCT04505345

(J Med Internet Res 2021;23(1):e18482) doi:10.2196/18482

KEYWORDS

alcohol use disorder; cognitive training; virtual reality

Introduction

Alcohol is a psychoactive substance that acts on the central nervous system, leading to dependence while causing severe physical, mental, and social problems [1]. Around 400 million people worldwide have problems related to excessive alcohol use, which is the third cause of death globally [2]. Alcohol abuse is also associated with diverse brain modifications [3], both at the structural and functional levels related to long-term, transient, or compensatory effects of alcohol [4], which have an impact on the integrity of the prefrontal brain cortex [5], causing deficits across a wide range of cognitive skills. These effects are particularly evident at the level of executive dysfunction, including attention, inhibitory control, behavioral control [3], verbal fluency, and decision making, similar to the effects of other substances, such as cannabis and cocaine [6]. Alcohol abuse in extreme cases may also lead to Korsakoff syndrome, which is characterized as an irreversible condition of anterograde amnesia [7].

These impairments at the brain level may disrupt behavior in such a way that individuals become overreactive to external cues related to the substance; this causes them to be unable to control substance-seeking behaviors and to make long-term decisions [8], which also affects treatment outcomes [9].

However, cognitive deficits resulting from chronic alcohol consumption are usually at least partially reversed during inpatient recovery periods, mostly as a result of abstinence [10]. While alcohol abuse is associated with morphological changes and reduced volume in multiple brain regions, mostly in the frontal lobe [5], abstinence from alcohol by patients recovering from alcohol use disorder (AUD) has been associated with significant recoveries in brain volume. The brain volume in these patients increases with the duration of the abstinence period; they also experience cognitive recoveries in speed and processing and shorter reaction times to stimuli [11]. Well-established treatments for alcohol dependence are, therefore, mostly designed with a focus on patients maintaining alcohol abstinence. Nevertheless, there is also agreement that the positive effects of abstinence at the cognitive level can be enhanced by cognitive interventions specifically designed to recover functions most compromised by alcohol abuse [12]; tests of the effectiveness of these interventions in the context of substance use disorders (SUDs) have been growing in recent years [13]. These cognitive interventions may be delivered in different formats. Cognitive stimulation typically involves group interventions for improvements not only in cognition but also socialization and does not focus on particular cognitive abilities. In contrast, cognitive training is a systematic training approach in which cognitive tasks are more focused on specific cognitive functions [14]. Its aim is to recover individuals' potential and minimize the impact of brain lesions via a set of programmed behavioral activities involving different cognitive functions [15]. For instance, Yohman and colleagues [16] found that patients who underwent classic cognitive training that focused on memory and problem-solving abilities showed greater improvements in the cognitive areas related to those abilities, but not to memory, than did control patients. Goldstein and colleagues [17] found that AUD participants exhibited significantly increased levels of perceptual and visuospatial skills, speed of information processing, and attention.

Meanwhile, recent technological progress has allowed new solutions for cognitive training based on the use of computerized systems to be developed. This trend is known as computerized cognitive training (CCT) and includes a growing number of systems available for cognitive training in different clinical contexts. An early study involving CCT did not show significant cognitive improvements from this treatment among individuals in residential treatment for AUD in comparison to controls [18], but more recent studies suggest that there are specific contributions from CCT for improving cognition in AUD patients. For instance, Fals-Stewart and Lam [12] found that individuals recovering from SUD, including abuse of alcohol, who underwent cognitive intervention within residential treatment programs reported better treatment outcomes than controls who underwent computer typing training and treatment as usual. Individuals in the experimental group reported being more involved both in the treatment and in the recovery program and stayed abstinent for longer periods of time; they also reported improved social and family behaviors. Two more recent studies using CCT have found specific cognitive improvements among patients in recovery programs for alcohol or other substances, with effects in executive functions [19], attention, delayed memory, and working memory [20].

A recent systematic review of cognitive training in AUD recovery suggests that these approaches may be useful to promote cognitive functioning on top of improvements due to mere abstinence. However, the available data with AUD individuals only provide evidence for near-transfer effects to very similar tasks, with no evidence regarding far-transfer effects to dissimilar tasks or to everyday functionality [13].

Some authors argue that virtual reality (VR)–based cognitive training is an especially ecologically valid form of CCT because it includes exercises that mirror everyday life activities and those that involve similar demands to those of everyday living [21]. For example, using VR-based serious games makes it possible to replicate different activities of daily living, such as tasks related to hygiene, having breakfast, choosing clothes, and going shopping, thus providing a function-led cognitive training approach [22], which may promote far-transfer effects of training to everyday living activities [23]. VR-based serious games have other advantages over traditional methods: immediate dynamic feedback, repeated practice, no physical consequences after an error, the setting, the fact that the task can be customized to the patient, and the fact that training involves progressive learning [19].

Thus, previous research suggests that there are positive effects of computer-based cognitive training, but the specific contributions of VR-based cognitive training reproducing everyday life activities have not yet been demonstrated. In particular, randomized controlled trials (RCTs) have been lacking. This study builds on previous research but proposes a pilot RCT to estimate the effects of a VR-based cognitive intervention on patients with AUD at the level of memory, attention functions, and executive functions; this approach will help in reducing biases associated with previous noncontrolled studies [13], while contributing to the estimation of effect sizes to determine necessary power for future definitive RCTs.

Methods

Trial Design

The study design consisted of a two-arm pre-post RCT in which participants were assigned either to an experimental group that underwent VR-based cognitive training combined with treatment as usual in residential community rehabilitation or to a passive control group comprising treatment as usual but without VR-based cognitive training. This trial consisted of an open-label RCT in which patients, researchers, assessors, and therapists were not blinded to group allocation. The allocation ratio was 1:1. Both groups underwent the same treatment program for alcohol recovery that is administered to all inpatients in residential treatment. Random allocation was concealed prior to the start of the study and was based on simple randomization after baseline assessment with random number generation in Microsoft Excel, Office 365. The intervention model consisted of a parallel design. The trial was retrospectively registered at ClinicalTrials.gov (NCT04505345).

Recruitment

The sample was recruited from September 2017 to May 2018 at a clinic for recovery from AUD, *Casa de Saúde do Telhal*, in the Lisbon region of Portugal. Recruitment was conducted during the psychological appointment of each patient's first week of treatment. Treatment for recovery from AUD in this institution is based on a 4-week residential treatment program following the Minnesota Model. To be included in the study, participants had to be at least 18 years old, with normal or corrected-to-normal vision and hearing, without a history of psychiatric or neurological disorders, and attending the inpatient program at the institution where the data were collected. Exclusion criteria were the presence of a psychosis episode during the program and withdrawal from the inpatient program. These criteria were checked by the clinician who conducted patient recruitment.

Outcomes

The outcomes of this study were selected from well-established neuropsychological tests. Primary outcomes were based on general cognitive functioning and executive functions, while secondary outcomes were based on specific cognitive tests for memory, attention, and cognitive flexibility.

Global Cognition

The Montreal Cognitive Assessment (MoCA) [24] was developed as a rapid tool for cognitive screening and has been validated for the Portuguese population [25]. The average duration of assessment completion ranges from 10 to 15 minutes, with no specified time limit. It consists of 11 items assessing cognitive domains, such as visuospatial orientation, naming, memory, attention, language, abstraction, evocation, and orientation. The maximum score is 30 points, with the cutoff point at 24 points.

Executive Functions

The Frontal Assessment Battery (FAB) [26] is a neurocognitive evaluation tool used to assess executive functions and has also been validated for the Portuguese population [27]. This test assesses six constructs: conceptualization, mental flexibility, programming, sensitivity to interference, inhibitory control, and environmental autonomy. The maximum score is 18 points, with three cutoff points: less than 12 indicates dementia, 12 to 14 indicates dysfunction, and 15 to 18 is in the normative range.

Memory

The Rey Complex Figure test (RCF) [28] was used to evaluate the capacity of motor perceptual organization, attention, and immediate nonverbal visual memory. This test is divided in two parts: the first part consists of an exercise in visual reproduction, followed by an interval of about three minutes, after which participants are asked to reproduce the same figure by memory (ie, the second part). The maximum total score of both parts is 36 points. Evaluation of the task is both quantitative (ie, performance time and points) and qualitative (ie, assessment of the level of reproduction). In this study we used only the quantitative scoring.

Attention

The Toulouse Pierón test (TP) [29] was used to evaluate permanent voluntary attention, concentration, resistance to fatigue, and stimulus processing. The test has a timed duration of exactly 10 minutes. During this time the participants have to identify the largest number of characters in a proof sheet from those that are indicated at the top of the sheet. The test measures the total number of characters processed (ie, hits) and omissions. The outcomes consist of correct responses, errors, and omissions along with dispersion index (DI) and working efficiency, which are given as follows:

$$\text{Dispersion index} = (\text{errors} + \text{omissions}) / \text{hits} \times 100$$

$$\text{Working efficiency} = \text{hits} - (\text{errors} + \text{omissions})$$

Cognitive Flexibility

The Wisconsin Card Sorting Test (WCST), developed by Grant and Berg [30] and studied by Nyhus and Barcelo [31], assesses strategic planning, abstract thinking, and capacity for perseveration and conceptualization. It also assesses the ability to use environmental feedback in modeling cognitive functioning, object-directed behaviors, and modeling of response impulsiveness. The test consists of 128 cards divided into two decks of 64 each. The cards vary in color (ie, red, green, yellow, and blue), shape (ie, triangles, stars, crosses, and circles), and number (ie, one, two, three, and four). Participants must match the 128 cards to the 4 stimulus cards that the examiner places on the table. The WCST was scored on six performance parameters: number of trials administered, number of total errors, number of perseverative errors, number of trials to complete first category, failure to maintain set, and conceptual-level responses.

Procedure

Overview

The ethics committee of the School of Psychology and Life Sciences of the host institution approved the human subjects protocol used in this study. It was conducted according to best practices on human research and the principles of the Declaration of Helsinki.

After reading and signing the offline informed consent document, participants from both groups first completed a sociodemographic and clinical data questionnaire and then a flexible neuropsychological assessment battery consisting of tests of global cognition (ie, MoCA), executive functions (ie, FAB), attention (ie, TP), memory (ie, RCF), and cognitive flexibility (ie, WCST). Participants took between 1 and 1.5 hours to complete this session. Participants were given a code to ensure anonymity, and the questionnaires and battery of tests were identified with this code to pair them with posttreatment assessments.

Participants in the experimental group underwent the VR intervention plus treatment as usual during their second and third weeks of hospitalization. This intervention consisted of 10 sessions, each lasting 30 to 40 minutes, which ran twice a week over a period of 5 weeks; these sessions consisted of the performance of VR exercises based on activities of everyday life guided by one therapist in individual sessions, increasing in difficulty from session to session. Participants in the control group underwent treatment as usual for AUD.

The training intensity can be considered low in terms of the number of sessions, session duration, and total treatment dose, which fell below the common range of 10 to 14 hours [13]. These sessions were conducted at *Casa de Saúde do Telhal* during the patients' psychological appointments for AUD recovery.

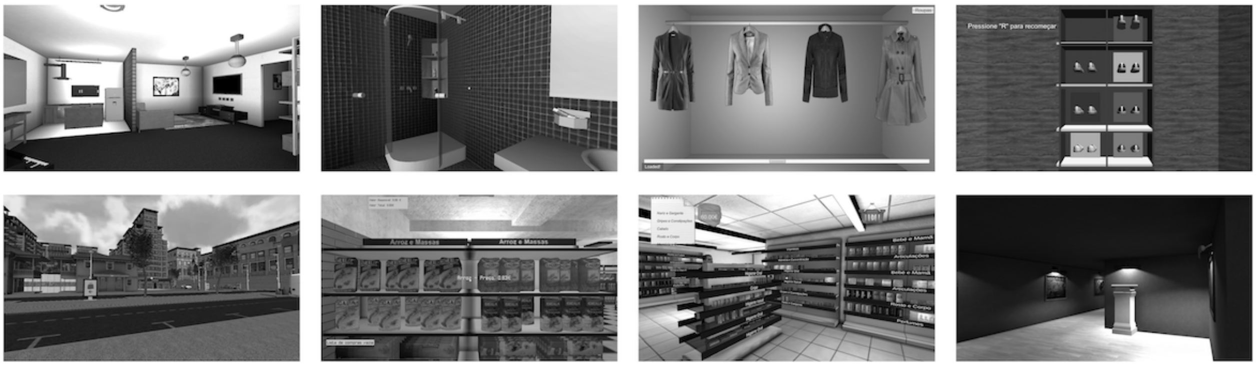
In the last week of hospitalization, both the control group and the experimental group participants completed the same battery of neuropsychological evaluation tests that were applied in the first session. This session was led by the same evaluator of the first assessment. None of the therapists or evaluators involved in these sessions were the owners of the software used.

Systemic Lisbon Battery

Cognitive training was conducted with the Systemic Lisbon Battery (SLB), release 2016 [19], a VR-based serious games platform that has been tested for cognitive training and assessment developed with Unity3D (Unity Technologies) for the Windows 10 system (Microsoft). Prior studies with the SLB for cognitive assessment have provided normative data with subsamples from the Portuguese population [19]. Regarding cognitive training, the SLB has been tested among different samples of participants with cognitive deficits due to brain injuries [32] and was also adapted to different technological platforms, such as mobile technology [33].

The SLB was developed as an alternative to the conventional methods of neurocognitive rehabilitation, but in this study it was used only for cognitive training. The platform consists of a virtual city with several built-up areas, a mini market, a pharmacy, an art gallery, and an interactive home; it also includes nonplayer characters walking around the city. The user is free to walk around the city and is given tasks to pick up certain objects in order to achieve a number of preset goals (eg, buy ingredients from a list at a grocery store). Elements from serious games, such as amount of money used or saved being used as a performance score along with visual and auditory feedback following completion of the tasks, are included in this platform to increase patients' motivation and retention rates with therapy. Figure 1 depicts some of the tasks comprising the SLB.

Figure 1. Examples of the virtual reality tasks used for cognitive training.



Statistical Analysis

The data were analyzed using SPSS Statistics for Windows, version 21.0 (IBM Corp). Normality was assessed by analyzing the distributions for each outcome according to skewness and kurtosis and was tested with the Shapiro-Wilk test. Skewness and kurtosis were within limits (±2) for the study variables, except for the DI of the TP. Shapiro-Wilk tests revealed that only the DI of the TP at both assessments and the copy trial from the RCF at postassessment violated the normality assumption. Therefore, those variables were assessed with nonparametric tests.

The baseline characteristics of the groups were compared using Student *t* tests for independent samples for outcomes with normal distribution and Mann-Whitney tests for independent samples for outcomes without normal distribution.

To test the effects of the treatment, we used repeated-measures analyses of covariance (ANCOVAs) with one within-subjects factor (ie, pretreatment vs posttreatment assessment) and one between-subjects factor (ie, experimental vs control group) while controlling for potential confounders. Effect sizes in the ANCOVAs are reported as η^2 and are given by the following equation:

$$\eta^2 = SSB / SST$$

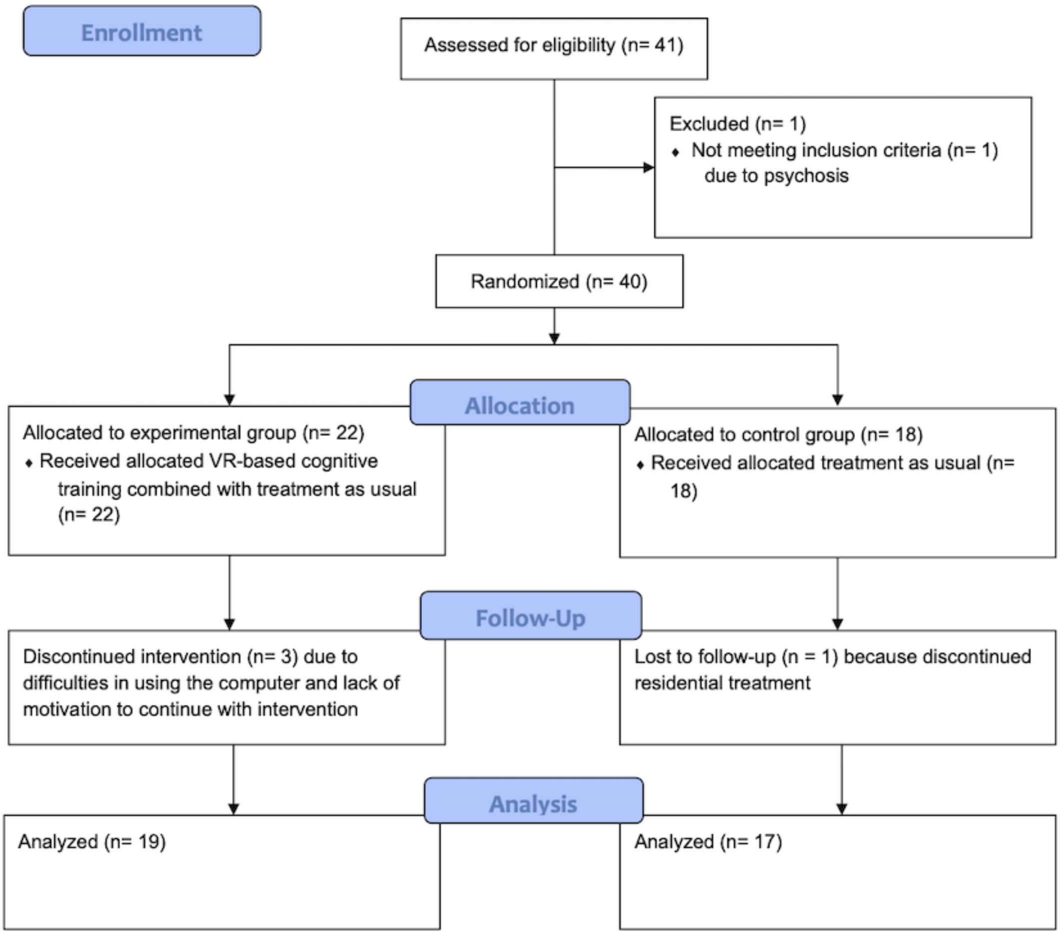
where SSB is the sum of squares for between-subjects factors and SST is the sum of squares for the total model. Following Cohen [34], we consider a small effect size as starting from 0.01, medium as starting from 0.06, and large as starting from 0.14, so as to inform future definitive RCTs. The statistical tests were explored with post hoc tests using the Bonferroni correction method. The statistical results were tested with an analysis of simple effects using Bonferroni correction. For nonnormal distributions, Wilcoxon signed-rank tests for two related samples were conducted separately for the experimental group and the control group. The significance level was set at α=.05.

Results

User Statistics

The final sample consisted of 36 patients; 30 (83%) were male, they were aged between 24 and 65 years (mean 44.83, SD 12.04), and the mean number of years they consumed alcohol was 14.31 (SD 4.34). From those 36 patients, 19 (53%) were assigned to the experimental group and 17 (47%) were assigned to the control group. From the total sample of 36 participants, 5 (14%) reported having used other substances in addition to alcohol in the past. Out of 36 participants, 6 (17%) had completed only 4 years of school (ie, elementary school), 22 (61%) had completed 6 years of school (ie, middle school), 7 (17%) had completed 9 years of school (ie, high school), and 1 (3%) had completed more than 12 years of school (ie, a higher degree). No differences were found between the experimental and control groups regarding gender distribution (*P*=.29) or level of education (*P*=.38). However, age did differ between the experimental and control groups; it was significantly higher in the control group (mean 51 years, SD 12.34) than in the experimental group (mean 40 years, SD 9.21; *P*=.004). Years consuming alcohol did not correlate with neuropsychological performance in these tests. The initial pool consisted of 41 patients who were recruited for the study, but 1 (3%) was excluded based on the exclusion criteria, while 4 (11%) were lost to follow-up: 3 from the experimental group were lost due to difficulties in using the computer and a lack of motivation to continue with cognitive training and 1 from the control group had discontinued residential treatment (see flow diagram in Figure 2).

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) 2010 flow diagram. VR: virtual reality.



Evaluation Outcomes

Baseline Comparisons Between Groups for Each Outcome

The comparisons between groups at the baseline assessment for each of the outcomes showed statistically significant differences in the total scores on the MoCA test (t_{34}=3.049; P=.004) and the FAB test (t_{34}=2.587; P=.01), suggesting better cognitive and executive functioning in the experimental group than in the control group.

However, because these differences could be due to the higher average age of controls, further analyses accounted for this by using the MoCA and FAB scores at baseline as covariates. Age did not correlate with MoCA (P=.62) or FAB scores (P=.58).

Pre-Post Comparisons for Each Outcome

Pre-post comparisons were performed using ANCOVAs with treatment assessment point (ie, time 1 vs time 2) as a within-subjects factor and treatment (ie, experimental vs control group) as a between-subjects factor controlling for baseline MoCA and FAB total scores on outcomes that were normally distributed. For outcomes that violated normal distribution, separate Wilcoxon tests were conducted for the experimental and control groups controlling these confounders. These analyses revealed a significant effect of the intervention on attention in two out of five of the TP outcomes and on cognitive flexibility in two out of six of the WCST outcomes. Improvements between pre- and posttreatment assessments in the experimental group were found for attention, concentration, cognitive flexibility, visual perception, and memory; these are discussed in more detail in the following three sections.

Attention and Concentration

The ANCOVA revealed a significant interaction effect between factors for correct responses ($F_{1,32}$=19.512; η^2=0.609; P<.001) and for working efficiency ($F_{1,32}$=10.986; η^2=0.343; P=.002). Simple effects (ie, Bonferroni corrected) showed a more pronounced improvement in both outcomes for the experimental group (P<.001). The same analysis comparing groups at each assessment point showed a difference between groups only at the posttreatment assessment for correct responses (P=.02) and for working efficiency (P=.03), as depicted in Table 1.

Table 1. Attention outcomes through the Toulouse Pierón test (TP).

Attention outcome	Experimental group score, mean (SE)		Control group score, mean (SE)	
	Pretreatment	Posttreatment	Pretreatment	Posttreatment
TP no correct responses	128.47 (7.95)	192.76 (10.12)	140.17 (8.48)	155.85 (10.79)
TP working efficiency	106.54 (8.40)	171.34 (17.57)	112.79 (8.96)	108.61 (18.73)

Wilcoxon tests were conducted separately for the experimental and control groups, controlling for MoCA and FAB scores that were divided into two groups by median split. Therefore, eight Wilcoxon tests were conducted comparing pre-post assessments for the DI: experimental group MoCA score of 23 and below (ie, median) versus experimental group MoCA score above 23, and experimental group FAB score of 14 and below (ie, median) versus experimental group FAB score above 14; the same design was used for the control group. These comparisons showed significant differences between pre-post assessments only for the group below the median score for the FAB, revealing a decrease in the DI for both the experimental group (*Z*=–2.240; *P*=.03) and the control group (*Z*=–2.934; *P*=.003), thus suggesting no group effects in pre-post assessments of the DI.

Cognitive Flexibility

The ANCOVA for each of the WCST outcomes indicated a significant interaction effect for the total number of errors ($F_{1,28}$=12.482; η^2=0.446; *P*=.001) and the number of trials to complete the first category ($F_{1,28}$=4.653; η^2=0.166; *P*=.04). Simple effects (ie, Bonferroni corrected) suggested more pronounced improvements in these outcomes in the experimental group (*P*<.001). A difference between groups was observed at the pretreatment assessment for the number of trials to complete the first category (*P*=.02), while at the posttreatment assessment a difference was found for the total number of errors (*P*<.001). A main effect was also observed for the number of trials administered ($F_{1,28}$=9.041; η^2=0.244; *P*=.006), indicating an improvement from pre- to posttreatment assessment in both groups. These results are shown in Table 2.

Table 2. Cognitive flexibility outcomes through the Wisconsin Card Sorting Test (WCST).

Cognitive flexibility outcome	Experimental group score, mean (SE)		Control group score, mean (SE)	
	Pretreatment	Posttreatment	Pretreatment	Posttreatment
WCST no errors	74.72 (3.14)	52.43 (2.70)	71.79 (3.62)	72.45 (3.12)
WCST no trials to complete first category	40.31 (3.92)	15.86 (1.62)	23.95 (4.53)	12.89 (1.87)

Visual Perception and Memory

Despite the fact that the ANCOVA did not reveal a significant effect in the total score from the RCF for the memory trial, the analysis of simple effects (ie, Bonferroni corrected) showed a significant improvement between pre- and posttreatment assessments in memory ability for the experimental group (*P*=.002) (see Table 3). The copy trial from the RCF was assessed using Wilcoxon tests conducted separately for the experimental and control groups, controlling for the MoCA and FAB scores: experimental group MoCA score of 23 and below (ie, median) versus experimental group MoCA score above 23, and experimental group FAB score of 14 and below (ie, median) versus experimental group FAB score above 14; the same design was used for the control group. Only one comparison was significant in the control group with FAB scores below the median (*Z*=–2.684; *P*=.007), which indicates that there were no consistent intervention effects on this outcome.

Table 3. Memory outcomes through the Rey Complex Figure test (RCF).

Memory outcome	Experimental group score, mean (SE)		Control group score, mean (SE)	
	Pretreatment	Posttreatment	Pretreatment	Posttreatment
RCF memory trial	21.54 (4.92)	37.80 (5.68)	22.16 (5.24)	24.99 (6.06)

Discussion

Principal Findings

There is evidence that cognition improves during recovery treatment for AUD, while such improvements may be enhanced by the specific effects of cognitive training [10]. Moreover, the use of VR tasks for CCT is better suited to promote the transfer of skills to everyday living, as these tasks are closer to real-life activities than most of the exercises used in classic cognitive training [23]. The aim of this study was to determine the specific contributions to cognition of a multidomain CCT using VR among residential patients recovering from AUD. For that, a pilot RCT was carried out to study the feasibility of this approach in distinguishing the effects of treatment with CCT. RCTs are the emerging gold standard for unbiased testing of treatments and have been lacking from this area of research.

The results found in this study converge with previous research [19,20] in which patients with AUD who underwent cognitive training showed greater cognitive improvements, particularly in areas related to executive functions, than those who did not undergo any particular cognitive treatment. In this study, specific attention effects and cognitive flexibility effects (ie, executive functions) were found.

At the initial assessment, the two groups had different results in the MoCA and FAB measures, suggesting different levels of general cognitive ability and executive functions. In order to control for the effect of this difference on the outcomes, these variables were controlled in the pre-post analyses. In those analyses, we found significant differences in cognitive performance between pre- and posttreatment for both groups in attention, as assessed with the TP, and in cognitive flexibility, a component of executive functions, as assessed with the WCST; this shows evidence of the positive effects of the residential treatment plan and alcohol abstinence, confirming a robust finding in the literature [10,11]. In addition, it was found that effect sizes in these outcomes were large and were more pronounced in attention outcomes than in cognitive flexibility.

However, in-depth analyses show that there were significant differences between the groups at the final assessment point for the correct responses and DI of the TP and total number of errors in the WCST; however, in other indicators (ie, number of trials to complete the first category in the WCST) that showed improvements between pre- and postassessments, the experimental and control groups did not significantly differ. This pattern of results suggests that there were effects by the VR training sessions in the experimental group beyond simple abstinence-related effects; this also suggests that the task-specific learning in cognitive training promoted neuroplasticity more clearly than abstinence by itself, thus promoting more consistent improvements in attention and cognitive flexibility, as specific components of executive functioning. In fact, executive dysfunctions typically associated with the prefrontal cortex are among the most pronounced deficits due to alcohol abuse [5].

Patient retention was another positive feature of this intervention that speaks to its feasibility, as only 14% (3/22) of patients discontinued the cognitive training intervention, which is near the lower bound of the range of 8% to 41% found in a recent review of previous studies [13].

Overall, this study suggests that the use of VR scenarios is a feasible option to enhance cognitive recovery in patients with AUD, specifically at the level of attention, and that it may also support improvements in cognitive flexibility; these are important cognitive abilities also underlying decision making and retention in recovery programs for AUD [35].

Limitations

One main limitation of this study was the small dose of the intervention. Given the time of stay for rehabilitation in residential treatment at the partner institution, it was not possible to extend rehabilitation over a longer intervention period. Thus, training intensity was inferior to most studies with cognitive rehabilitation [13], which may have compromised the ability of this trial to detect other positive intervention outcomes, enhancing a type II error. The small sample size and low power also enhanced a possible type II error, and replication with larger samples is still needed. On the other hand, the fact that this study was not double-blinded could have increased a type I error. The use of a passive control group consisting of treatment as usual, but without virtual training, precluded the use of blinding procedures to both patients and evaluators. It is important that further studies have an independent evaluator blinded to patient assignment.

Conclusions

In this study, we found a positive impact of the VR training on the cognitive rehabilitation, particularly on attention and executive functions, of individuals with AUD. Although the residential treatment according to the Minnesota Model has as its main objective the promotion and maintenance of abstinence behavior in relation to alcohol, it can also promote a recovery from alcohol dependence. Such cognitive improvements may not only contribute to a better quality of life among patients but also to their social and family functioning and, therefore, their ability to maintain abstinence.

Future studies should focus on more general outcomes related to functionality, well-being, or quality of life to help understand whether such cognitive-focused approaches also contribute to overall psychological adjustment or whether there are far-transfer effects of skills to dissimilar tasks than those trained in the program. It is also worth studying whether these effects remain stable with time when assessed at longer follow-ups.

Acknowledgments

PG was responsible for study conceptualization, writing the original draft, methodology, and software development and is the owner of the software used in this study. JO was also responsible for study conceptualization, conducted the formal analysis, and wrote, reviewed, and edited the manuscript. MM and EC, under the supervision of AD and PL, prepared the evaluation protocol and conducted the data collection. PL was also involved in study conceptualization. RB was responsible for revising and editing the final manuscript. All authors contributed to, and have approved, the final manuscript.

Conflicts of Interest

None declared.

References

1. Savic M, Room R, Mugavin J, Pennay A, Livingston M. Defining "drinking culture": A critical review of its meaning and connotation in social research on alcohol problems. Drugs 2016 Apr 15;23(4):270-282. [doi: 10.3109/09687637.2016.1153602]
2. Management of Substance Abuse Unit, Department of Mental Health and Substance Abuse, World Health Organization. Global Status Report on Alcohol and Health 2014. Geneva, Switzerland: World Health Organization; 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/112736/9789240692763_eng.pdf?sequence=1 [accessed 2020-01-02]
3. Oscar-Berman M, Valmas M, Sawyer K, Ruiz S, Luhar R, Gravitz ZR. Profiles of impaired, spared, and recovered neuropsychologic processes in alcoholism. Handb Clin Neurol 2014;125:183-210 [FREE Full text] [doi: 10.1016/B978-0-444-62619-6.00012-4] [Medline: 25307576]
4. Sullivan EV, Pfefferbaum A. Neurocircuitry in alcoholism: A substrate of disruption and repair. Psychopharmacology (Berl) 2005 Aug;180(4):583-594. [doi: 10.1007/s00213-005-2267-6] [Medline: 15834536]
5. Moselhy HF, Georgiou G, Kahn A. Frontal lobe changes in alcoholism: A review of the literature. Alcohol Alcohol 2001;36(5):357-368. [doi: 10.1093/alcalc/36.5.357] [Medline: 11524299]
6. Fernández-Serrano MJ, Pérez-García M, Schmidt Río-Valle J, Verdejo-García A. Neuropsychological consequences of alcohol and drug abuse on different components of executive functions. J Psychopharmacol 2010 Sep;24(9):1317-1332. [doi: 10.1177/0269881109349841] [Medline: 20007413]
7. Zahr NM, Kaufman KL, Harper CG. Clinical and pathological features of alcohol-related brain damage. Nat Rev Neurol 2011 May;7(5):284-294. [doi: 10.1038/nrneurol.2011.42] [Medline: 21487421]
8. Noël X, Brevers D, Bechara A. A neurocognitive approach to understanding the neurobiology of addiction. Curr Opin Neurobiol 2013 Aug;23(4):632-638 [FREE Full text] [doi: 10.1016/j.conb.2013.01.018] [Medline: 23395462]
9. Fals-Stewart W, Schafer J, Lucente S, Rustine T, Brown L. Neurobehavioral consequences of prolonged alcohol and substance abuse: A review of findings and treatment implications. Clin Psychol Rev 1994 Jan;14(8):755-778. [doi: 10.1016/0272-7358(94)90041-8]
10. Manning V, Wanigaratne S, Best D, Hill RG, Reed LJ, Ball D, et al. Changes in neuropsychological functioning during alcohol detoxification. Eur Addict Res 2008;14(4):226-233. [doi: 10.1159/000156479] [Medline: 18810242]
11. Durazzo TC, Mon A, Gazdzinski S, Meyerhoff DJ. Regional brain volume changes in alcohol-dependent individuals during early abstinence: Associations with relapse following treatment. Addict Biol 2017 Sep;22(5):1416-1425 [FREE Full text] [doi: 10.1111/adb.12420] [Medline: 27329647]
12. Fals-Stewart W, Lam WKK. Computer-assisted cognitive rehabilitation for the treatment of patients with substance use disorders: A randomized clinical trial. Exp Clin Psychopharmacol 2010 Feb;18(1):87-98. [doi: 10.1037/a0018058] [Medline: 20158298]
13. Nixon SJ, Lewis B. Cognitive training as a component of treatment of alcohol use disorder: A review. Neuropsychology 2019 Sep;33(6):822-841. [doi: 10.1037/neu0000575] [Medline: 31448949]
14. Clare L, Woods RT. Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: A review. Neuropsychol Rehabil 2004 Sep;14(4):385-401. [doi: 10.1080/09602010443000074]
15. Wilson BA, Gracey F, Evans JJ, Bateman A. Neuropsychological Rehabilitation: Theory, Models, Therapy and Outcome. Cambridge, UK: Cambridge University Press; 2009.
16. Yohman JR, Schaeffer KW, Parsons OA. Cognitive training in alcoholic men. J Consult Clin Psychol 1988 Feb;56(1):67-72. [doi: 10.1037//0022-006x.56.1.67] [Medline: 3346451]
17. Goldstein G, Haas GL, Shemansky WJ, Barnett B, Salmon-Cox S. Rehabilitation during alcohol detoxication in comorbid neuropsychiatric patients. J Rehabil Res Dev 2005;42(2):225-234 [FREE Full text] [doi: 10.1682/jrrd.2004.03.0040] [Medline: 15944887]
18. Peterson MA, Patterson B, Pillman BM, Battista MA. Cognitive recovery following alcohol detoxification: A computerised remediation study. Neuropsychol Rehabil 2002 Jan;12(1):63-74. [doi: 10.1080/09602010143000167]
19. Gamito P, Oliveira J, Brito R, Lopes P, Rodelo L, Pinto L, et al. Evaluation of cognitive functions through the Systemic Lisbon Battery: Normative data. Methods Inf Med 2016;55(1):93-97. [doi: 10.3414/ME14-02-0021] [Medline: 26660359]
20. Rupp CI, Kemmler G, Kurz M, Hinterhuber H, Fleischhacker WW. Cognitive remediation therapy during treatment for alcohol dependence. J Stud Alcohol Drugs 2012 Jul;73(4):625-634. [doi: 10.15288/jsad.2012.73.625] [Medline: 22630801]
21. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Human Behav 2017 May;70:317-327. [doi: 10.1016/j.chb.2017.01.013]
22. Parsons TD. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci 2015;9:660 [FREE Full text] [doi: 10.3389/fnhum.2015.00660] [Medline: 26696869]
23. Doniger GM, Beeri MS, Bahar-Fuchs A, Gottlieb A, Tkachov A, Kenan H, et al. Virtual reality-based cognitive-motor training for middle-aged adults at high Alzheimer's disease risk: A randomized controlled trial. Alzheimers Dement (N Y) 2018;4:118-129 [FREE Full text] [doi: 10.1016/j.trci.2018.02.005] [Medline: 29955655]
24. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]

25. Freitas S, Simões MR, Alves L, Santana I. Montreal Cognitive Assessment (MoCA): Normative study for the Portuguese population. J Clin Exp Neuropsychol 2011 Nov;33(9):989-996. [doi: 10.1080/13803395.2011.589374] [Medline: 22082082]
26. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: A Frontal Assessment Battery at bedside. Neurology 2000 Dec 12;55(11):1621-1626. [Medline: 11113214]
27. Lima CF, Meireles LP, Fonseca R, Castro SL, Garrett C. The Frontal Assessment Battery (FAB) in Parkinson's disease and correlations with formal measures of executive functioning. J Neurol 2008 Nov;255(11):1756-1761. [doi: 10.1007/s00415-008-0024-6] [Medline: 18821046]
28. Rey A. L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems). Arch Psychol (Geneve) 1941;28:215-285.
29. Toulouse E, Piéron H. Toulouse-Piéron: Prueba Perceptiva y de Atención. Manual. Madrid, Spain: TEA Ediciones; 1986.
30. Grant DA, Berg E. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. J Exp Psychol 1948 Aug;38(4):404-411. [doi: 10.1037/h0059831] [Medline: 18874598]
31. Nyhus E, Barceló F. The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update. Brain Cogn 2009 Dec;71(3):437-451. [doi: 10.1016/j.bandc.2009.03.005] [Medline: 19375839]
32. Oliveira J, Gamito P, Lopes B, Silva AR, Galhordas J, Pereira E, et al. Computerized cognitive training using virtual reality on everyday life activities for patients recovering from stroke. Disabil Rehabil Assist Technol 2020 Apr 07:1-6. [doi: 10.1080/17483107.2020.1749891] [Medline: 32255695]
33. Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Caçoete C, et al. Cognitive training through mHealth for individuals with substance use disorder. Methods Inf Med 2017 Mar 23;56(2):156-161. [doi: 10.3414/ME16-02-0012] [Medline: 28116414]
34. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
35. Shulman M, Campbell A, Pavlicova M, Hu M, Aharonovich E, Nunes EV. Cognitive functioning and treatment outcomes in a randomized controlled trial of internet-delivered drug and alcohol treatment. Am J Addict 2018 Sep;27(6):509-515 [FREE Full text] [doi: 10.1111/ajad.12769] [Medline: 30091814]

Abbreviations

ANCOVA: analysis of covariance
AUD: alcohol use disorder
CCT: computerized cognitive training
DI: dispersion index
FAB: Frontal Assessment Battery
MoCA: Montreal Cognitive Assessment
RCF: Rey Complex Figure test
RCT: randomized controlled trial
SLB: Systemic Lisbon Battery
SSB: sum of squares for between-subjects factors
SST: sum of squares for the total model
SUD: substance use disorder
TP: Toulouse Pierón test
VR: virtual reality
WCST: Wisconsin Card Sorting Test

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by F Tozzi, J Vermeir; comments to author 12.06.20; revised version received 10.08.20; accepted 11.11.20; published 29.01.21.

Please cite as:
Gamito P, Oliveira J, Matias M, Cunha E, Brito R, Lopes PF, Deus A
Virtual Reality Cognitive Training Among Individuals With Alcohol Use Disorder Undergoing Residential Treatment: Pilot Randomized Controlled Trial
J Med Internet Res 2021;23(1):e18482
URL: http://www.jmir.org/2021/1/e18482/
doi:10.2196/18482
PMID:33512329

Original Paper

Perceived Needs Versus Predisposing/Enabling Characteristics in Relation to Internet Cancer Information Seeking Among the US and Chinese Public: Comparative Survey Research

Di Zhang[1], PhD; Hongchao Hu[1], MA; Zhen Shi[1], MSc; Biao Li[1], PhD

[1] School of Journalism and Communication, Renmin University of China, Beijing, China

Corresponding Author:
Biao Li, PhD
School of Journalism and Communication
Renmin University of China
RM713 Mingde Journalism Buliding, Renmin University of China
59 Zhongguancun Rd, Haidian Dist
Beijing, 100872
China
Phone: 86 18810286586
Email: libiao@ruc.edu.cn

Abstract

Background: Since the rise of the internet, online health information seeking has become a worldwide phenomenon. However, health and health communication are inherently culture bound. A data-driven cross-country comparison enables us to better understand how cultural factors moderate the association between individual-level determinants and online health information seeking.

Objective: The objective of the study was to examine similarities and differences in determinants of internet cancer information seeking between the US and Chinese general public (excluding cancer patients and survivors) under the framework of a behavioral model of health services use.

Methods: This study used Health Information National Trends Survey (HINTS) 2017 (US data) and HINTS-China 2017 data to answer the research question. It focused on people with no cancer history and with internet access. For the HINTS 2017, the sample size was 2153; for the HINTS-China 2017, the sample size was 2358. To compare China and the United States, the researchers selected the same set of study variables for each dataset. Under the framework of the behavioral model of health services use, these predictors were predisposing factors, enabling factors, and need factors.

Results: In terms of the predisposing factors, a higher age, college degree or above, being currently unemployed, and having a family history of cancer were associated with internet cancer information seeking for the Chinese respondents; none of these factors were related to information seeking for the US respondents, although a lower age was associated with information seeking. Regarding the enabling conditions, lower trust in family members and friends as reliable information sources was the only factor associated with information seeking for the Chinese respondents, while no enabling factor was related to information seeking for the US respondents. Regarding the need factors, perceived health status was not related to information seeking for the Chinese respondents, while perception of poorer health condition was related to information seeking for the US respondents. Higher cancer fear was related to information seeking for both groups, but the magnitude of association was smaller for the Chinese respondents than for the US respondents.

Conclusions: Overall, under the framework of the behavioral model of health services use, the results based on multivariate logistic regression reveal clear patterns of cross-country/cultural differences in the factors associated with internet cancer information seeking behaviors: predisposing characteristics and enabling conditions are more important in China, while perceived needs are more significant in the US. Such differences might reflect possible US-China differences in job environment (eg, job pressure) and culture (individualism vs collectivism and family structure).

(J Med Internet Res 2021;23(1):e24733) doi:10.2196/24733

KEYWORDS

HINTS; health information seeking behavior (HISB); China; United States; comparative research; cultural sensitivity

Introduction

Background

A milestone in monitoring the US public's access to and use of health information is the Health Information National Trends Survey (HINTS) initiated by the National Cancer Institute [1], which provides valuable guidance for practitioners. With the cultural sensitivity approach becoming increasingly popular [2], calls for the expansion of the HINTS research programs into other parts of the world have been made, and HINTS-China pioneered this international expansion [3]. The pilot international HINTS survey has been conducted twice in China, in 2012 and 2017 [3]. The replication of the HINTS research program in countries and regions outside the US based on similar measurement schemes benefits global health communication research in numerous ways; one is that it allows scholars to compare health information seeking behaviors among countries [3]. Data-driven, multicountry comparisons enable us to better understand how cultural factors moderate the association between individual-level determinants and health information seeking, which further helps practitioners evaluate the probability of successful health communication intervention designs in one country being transferable to other countries [3].

To fulfill the mission of the international expansion of the HINTS program, this study used HINTS and HINTS-China data collected in 2017 [4] to examine similarities and differences in determinants of online cancer information seeking between the US and Chinese general public (excluding cancer patients and survivors) under the framework of the behavioral model of health services use [5,6]. Since the rise of the internet, online health information seeking has attracted worldwide scholarly attention. However, most scholars have conducted such studies in a single country. Some scholars have designed comparative studies across countries [7,8], but they have normally been based on small nonprobability college student samples. Therefore, the results of this study can contribute to the health communication literature by generating more reliable insights into country differences in health information acquisition, such as internet cancer information seeking.

Literature Review

Determinants of Online Health Information Seeking

We not only reviewed studies on cancer information seeking on the internet, which are limited in number, but also reviewed research concerning generalized online health information seeking to present a more complete picture of the factors that potentially play a role. Regardless of the theoretical models used, the predictors of major concern in existing studies include mainly demographics [9], structural characteristics [9], and perception variables [10]. Additionally, such studies often assess physiological indicators, but they have rarely been found to be statistically significant [11].

The effects of demographic and structural variables are rather stable. According to HINTS data, Americans who refer to the internet for health information are younger than those who do not [9,11]. Additionally, age moderates the positive association between trust in internet information and online health information seeking [11]. Previous research has generally found that higher socioeconomic status is associated with greater health information acquisition [9,12,13]. Married people reported a higher frequency of internet use to search for health information [12]. However, the results related to gender are mixed. Some studies have found that women are more likely to use the internet to obtain health information than men [12,14], while others have found that gender does not make a difference [9,15].

Most studies on generic health information seeking concentrate on people's perceptions of risk, personal health status, self-efficacy, media trust, social support, and satisfaction with caregivers. Those with higher levels of perceived risks and fears and lower levels of confidence about their health status are more likely to search for health information on the internet [10,16-18]. Those with higher internet self-efficacy are more likely to obtain health information online [10,19-21]. Those who trust the internet as an accurate and reliable information source or channel are more likely to rely on it to obtain medical information [11,18,19]. Additionally, trust in other players or information sources is positively associated with online health information seeking behaviors. Trust in health information from family members, for instance, is positively related to internet health information seeking behaviors among the US population [22]. Social support is positively associated with web searches for health information [21,22]. People with more social ties can seek health information from their social ties when finding it difficult to obtain online [23]. In a similar vein, patients who are unsatisfied with doctors tend to use the internet for medical information [24].

Behavioral Model of Health Services Use and Health Information Seeking Predictors

We used the behavioral model of health services as the theoretical framework to organize the predictor variables of internet cancer information seeking, as the model evaluates the extent to which 3 sets of predictor variables (predisposing factors, enabling factors, and need factors) influence people's use of health services [5,6,25]. The model also suggests a causal order: predisposing characteristics such as demographics and social structures are exogenous; enabling factors are necessary conditions for people to use medical services; and need variables (perceptions of needs) must be defined for use to take place [25]. According to the model, the 3 categories of factors encompass both contextual and individual characteristics [26]. Contextual characteristics refer to the environment and circumstances of health care service access, health care provider characteristics, and community characteristics, which are often measured at the aggregate level instead of the individual level [26]. Individual characteristics include demographics, social factors, personal beliefs, personal financial status, personal perceptions of health status, and objective evaluation of one's physical status [26]. Contextual characteristics may influence access to health care services through individual characteristics [26]. Although the behavioral model of health services was originally proposed to explain people's use of health services, scholars have successfully applied it to explain variations in health information seeking or channel use [27,28]. Additionally, the model incorporating both contextual and individual characteristics is suitable for the comparative nature of this study because the US and China differ in terms of community structure, culture, and health traditions [29-32].

This study focused on how individual-level predisposing, enabling, and need factors are associated with information seeking and how their relationships vary as a function of the country group. Predisposing characteristics were age; gender; education; ethnicity; the status of a person in his or her community; general health attitudes, values, and knowledge; and genetic characteristics [25,26]. Enabling resources were access to health facilities, access to medical personnel, and access to the medical system, including health care policies, personal income, health insurance, social relationships, and organizational factors [25,26]. Need factors were an individual's evaluation of his or her health status, need for medical services, and concerns about health problems [25,26]. With reference to a study that analyzed HINTS-China 2012 data [27], the researchers of this study selected age, gender, educational attainment, marital status, and BMI as predisposing factors. Enabling factors were current occupational status, household income, and trust in social institutions and information channels. Need characteristics were perceived health status, fatalism about cancer, fear of cancer, and perceived cancer risks. Although these variables were measured at the individual level, Anderson's model suggests that contextual characteristics can exert an influence through individual-level characteristics, which can reflect meso- or macrolevel contextual influences [29-32]. For instance, occupational status and household income can reflect the influence of the mesolevel working environment; trust in family members as health information sources can reflect mesolevel family influences; and perceived need factors can reflect macrolevel cultural influences.

Differences Between the US and China

Health communication scholars have paid increasing attention to internet health information seeking in the US and Europe, but little is known about China, where more than 4 million patients were diagnosed with cancer in 2015 [33]. Health and health communication are inherently culture bound [2]. Research findings generated from the West may not be readily applicable to countries such as China and Japan, which differ significantly from the US and other Western countries. Because this study compares China and the US, the following sections focus on cultural and structural differences between the two countries.

China and the US vary in issues related to cancer control. First, the types of common cancers in China differ from those in the US. In China, lung cancer has the highest incidence rate [34]. In the US, prostate cancer has the highest incidence rates for men and breast cancer for women [35]. Second, the Chinese medical care system is not as well established as the US medical care system [36]. With market-oriented reforms, the traditional system of the government and state-owned enterprises taking care of all medical needs has collapsed, while the emerging medical insurance system is far from satisfactory [36,37]. Third, many Chinese people believe in Chinese traditional medicine based on a holistic worldview [32]. Some Chinese cancer patients choose integrated therapies combining Western and Chinese traditional medicines [38]. Fourth, the doctor-patient relationship in China has been deteriorating for more than a decade [39]. Unlike their US counterparts, Chinese families do not have family doctors and, as such, visit comprehensive and specialized public hospitals even for minor conditions such as colds and coughs; as a result, a huge workload makes it difficult for doctors and nurses in China to find time to communicate with patients and their family members [39]. Workplace violence involving doctors and patients/family members has often made Chinese media headlines [40]. The commercialized medical care system in China has further eroded public trust in doctors, who people believe conspire with pharmaceutical companies to make money off of patients [41].

Chinese culture differs from US culture. China is collectivistic, while the US is more individualistic [30]. Individualism versus collectivism is an important construct in health psychology; individuals in individualistic cultures tend to emphasize personal wellbeing [42]. Individualism versus collectivism moderates the relationship between negative affect and psychological wellbeing, and the strength of the negative association between negative affect and wellbeing is stronger in individualistic cultures than in collectivistic cultures [43]. Additionally, Chinese people are commonly believed to be more family oriented than their US counterparts [44]. In terms of social structure, although the one-child policy has destroyed the traditional Chinese extended family structure, families remain central in Chinese life [45]. Family members remain close to each other by staying in a so-called networked family, even though it is no longer characterized by physical proximity among members [45]. Comparing differences in the strength of family-related variables can partly address criticism that the behavioral model of health services use lacks consideration of social relationships [25].

Although the US has a longer history of internet commercialization, China has recently begun to take the lead in the commercialization of new media applications, which have penetrated every social stratum [46]. A case in point is that most street beggars print out QR codes to collect donations in the street. Therefore, Chinese people increasingly use various types of internet applications to obtain health information [47]. Additionally, China is ahead of the US in the delivery of internet medical services because venture capitalists and entrepreneurs have invested heavily in the internet medical care industry [48]. It is common for Chinese people to make medical appointments with doctors using internet applications such as hospital portals, social media, and mobile applications [47].

On the basis of the preceding discussion, this study plans to answer the following research question:

> *How do China and the US differ in the associations between predisposing/enabling/need factors and internet cancer information seeking behaviors among people without a history of cancer?*

Methods

Data Sources

This study used HINTS 2017 (US data) and HINTS-China 2017 data to answer the research question. HINTS 2017 had a final sample size of 3285. Of the respondents, 2756 had no history of cancer (current patients or survivors); thus, these respondents were the focus of the study.

HINTS-China is an effort jointly launched by the Chinese Ministry of Health Center for Health Education, Renmin University of China, the Chinese National Cancer Center, and George Mason University and was recently joined by Beijing Normal University [27]. HINTS-China modeled the majority of the question items in HINTS and was adapted to local Chinese characteristics and the Chinese population. In the pilot phase, the Chinese survey was conducted twice, in 2012 and 2017, and was administered only in the Chinese capital city, Beijing, and Anhui Province's capital city, Hefei. Beijing was chosen because it is a megacity and the economic center of China, and Hefei was selected because it is a less developed, second-tier Chinese city that represents a more typical Chinese city. HINTS-China 2017 followed the multistage stratified random sampling strategy of HINTS-China 2012 [25]. HINTS-China 2017 was conducted using in-person interviews, had a valid response rate of 64%, and resulted in a sample size of 3090, of whom 10 had a history of cancer.

Because online health information seeking is directly related to physical access to the internet, this study excluded those without internet access from the final model because it is obvious that those without internet access cannot search for cancer information on the web. The US sample contained 750 respondents who had never accessed the internet, and the Chinese sample had 723. Thus, the sample sizes of cancer-free populations with internet access were 2,153 and 2,358 for the US and China, respectively.

To compare China and the US, the researchers selected the same set of study variables as predictors in HINTS 2017 and HINTS-China 2017. Additionally, there were slight differences in the measurement scales of these predictors between countries, so recoding was conducted to facilitate comparison. Measurements of the predisposing, enabling, need, and outcome variables were as follows.

Variable Measurement

Predisposing Variables

Gender was recoded as female, with 1 = female and 0 = male. Educational attainment for the US and China was converted from 7-point and 6-point scales, respectively, to 4-point scales, with 1 = "Less than high school," 2 = "High school," 3 = "Vocational school," and 4 = "College and above." Marital status for both countries was recoded as "Currently married," with 1 = "Yes" and 0 = "No." Similarly, current occupational status was converted to "Currently employed," with 1= "Yes" and 0 = "No." Current smoking status was assessed by 2 items, which were the same in both countries' surveys. The first item was a filter question, "Have you smoked at least 100 cigarettes in your entire life?," with 2 choices, (1) "Yes" and (2) "No." Only those who answered "Yes" were presented with the second item, "How often do you now smoke cigarettes?," with 3 choices (1) "Every day," (2) "Some days," and (3) "Not at all." To combine the items into a single item, the researchers counted "No" for the first item as "Not at all" for the second item. BMI was calculated from height and weight using the standard formula. Both the Chinese and US surveys collected height and weight data. Family cancer history was measured using one item in both surveys. The US survey had 3 choices, (1) "Yes," (2) "No," and (3) "Not sure." The Chinese version broke "Yes" into "Close relatives" and "Distant relatives." They were recoded as a single variable, with 1 = "Yes" (including "Close relatives" and "Distant relatives") and 0 = "No or not sure."

Enabling Variables

Household annual income was originally measured at the interval level in China, which differs from the categorical income range used in HINTS 2017. Income in the Chinese version was recoded as a categorical variable in 4 quartiles (from "50,000 RMB yuan and below" to "150,001 RMB yuan and above" with a 50,000-RMB interval). Income in the US version, with 9 original categories, was also recoded in quartiles: "$0 to $19,999," "$20,000 to $49,999," "$50,000 to $99,999," and "$100,000 or more."

HINTS in both countries measured people's trust in social institutions and information channels as reliable health information sources. However, the Chinese and US surveys differed slightly in the design of the attributes of each question item. First, the US survey measured only people's trust in the generic internet, while the Chinese survey measured people's trust in 8 typical internet applications. Then, the researchers averaged the trust score for the 8 internet applications to create an overall score for trust in the internet for the Chinese data. Second, the US survey used a single item to measure people's trust in family or friends, while the Chinese survey measured them separately. Thus, the researchers averaged them to create a single score. Third, the US survey measured people's trust in newspapers or magazines, while the Chinese survey measured them separately. The researchers created a new item using the mean score. Fourth, the US survey measured trust using a 4-point scale, with 1 = "Not at all" and 4 = "A lot." In contrast, the Chinese survey measured trust using a 5-point scale, with 1 = "Very untrustworthy" and 5 = "Very trustworthy." The 4-point scale used in the US survey was converted numerically into a 5-point scale to make the comparison more straightforward. Taking into account the goal of the study, the researchers categorized trust variables into 5 groups. The first category was trust in social institutions, including government agencies, religious organizations, and charities, which were combined into an additive index. The second category was trust in traditional media channels, including print media, television, and radio, which were combined into an additive index (Cronbach alpha for US=.77; Cronbach alpha for China=.85). Trust in the internet, doctors, and family members and friends were used as they were in the following logistic regression analysis.

Need Variables

Self-confidence about personal health or perceived health status was measured on a 5-point scale, with 1 = Poor and 5 = Excellent. The scales for both countries were the same.

Both surveys measured fatalism about cancer, which was assessed by 4 items such as "It seems like everything causes cancer" and "When I think about cancer, I automatically think about death.” The US survey used a 4-point scale, with 1 = “Strongly agree” and 4 = “Strongly disagree.” The scale in the Chinese survey was slightly different, with 1 = “Strongly disagree” and 5 = “Strongly agree.” The scale in the US survey was reversed and numerically converted into a 5-point scale, thus allowing the researchers to more easily compare the magnitudes of the coefficients. They were added to an index (Cronbach alpha for US=.63; Cronbach alpha for China=.75).

Fear of cancer was assessed using a single item on a 5-point scale, with 1= “Not at all” and 5= “Extremely.” The surveys for the two countries were identical.

Perceived cancer risk was assessed by asking the respondents, “How likely are you to get cancer in your lifetime?” with 1 = “Very unlikely” and 5 = “Very likely.” The surveys for the two countries were identical.

Outcome Variable

Online cancer information seeking was the outcome variable. In HINTS 2017, online cancer information seeking was binary, with 1 = “Yes” and 0 = “No,” while HINTS-China 2017 measured it using a 4-point scale. To make them comparable, the researchers recoded “Never” and “Rarely” as “No,” and converted “Often” and “Sometimes” to “Yes.”

Data Analysis

In the first step, the researchers presented descriptive statistics of all the variables under examination. In the second step, the researchers used multivariate logistic regression to explore US-China differences in the associations between the 3 categories of the predictor variables and internet cancer information seeking (Table 1). The first model of the multivariate analysis examined only the main effects of the predictors; odds ratios, which are exponentiated log odds, were presented. The second model incorporated interaction terms of the country group (the US vs China) and predictor variables to examine between-country variation. The odds ratios of interaction terms are ratios of odds ratios. In the third step, the researchers presented conditional odds ratios of each predictor variable for each country group separately, and those predictor variables with statistically significant interaction terms were the focus of the discussion.

The US HINTS sample had missing values in most variables selected for comparison, and the listwise approach would have resulted in the loss of 569 observations (over 25% of the US sample cases). To maximize the number of observations used in the analysis, this study used the R package MICE (Multivariate Imputation via Chained Equations) to impute the missing values. Specifically, the researchers applied the CART (classification and regression trees) algorithm for categorical variables and PMM (predictive mean matching) for numerical variables [49]. A total of 50 sets of data with imputed values were generated for the US and China, and the coefficients and standard errors of the logistic regression results for the 50 imputed datasets were pooled together using the Rubin rules [50].

Table 1. Odds ratio of logistic regressions of predisposing, enabling, and need variables of internet cancer information seeking.

Variables	Model 1 odds ratio (95% CI)	Model 2 odds ratio (95% CI)
Predisposing variables		
Female (vs male)	1.07 (0.89-1.28)	1.11 (0.83-1.49)
Age	0.99 (0.99-1.00)*	1.02 (1.00-1.03)*
Education (vs less than high school)		
High school	0.88 (0.62-1.24)	1.08 (0.72-1.62)
Vocational school	1.23 (0.87-1.73)	1.58 (1.03-2.42)
College and above	1.53 (1.08-2.18)*	2.35 (1.50-3.68)*
Currently married (vs not)	1.15 (0.95-1.39)	0.97 (0.69-1.36)
Currently employed (vs not)	0.80 (0.66-0.98)*	0.66 (0.50-0.89)*
BMI	0.99 (0.97-1.01)	0.98 (0.94-1.03)
Family cancer history (vs not)	1.51 (1.25-1.82)*	2.23 (1.73-2.88)*
Smoking status (vs every day)		
Some days	1.04 (0.61-1.78)	0.54 (0.20-1.46)
Not at all	1.12 (0.85-1.48)	0.94 (0.64-1.39)
Enabling variables		
Annual family income (vs \$0 to \$19,999/50,000 RMB and below)		
\$20,000 to \$49,999 (50,001 to 100,000 RMB)	0.74 (0.58-0.94)*	0.70 (0.51-0.96)*
\$50,000 to \$99,999 (100,001 to 150,000 RMB)	0.65 (0.49-0.86)*	0.65 (0.43-0.99)*
\$100,000 or more (150,001 RMB and above)	0.76 (0.57-1.02)	0.71 (0.48-1.06)
Trust in information sources		
Social institutions	0.99 (0.95-1.04)	0.97 (0.90-1.04)
Traditional media	1.02 (0.98-1.07)	1.01 (0.95-1.08)
Internet	1.18 (1.05-1.32)*	1.43 (1.13-1.80)*
Doctors	0.93 (0.84-1.04)	0.96 (0.81-1.13)
Family and friends	0.87 (0.78-0.97)*	0.71 (0.59-0.86)*
Need variables		
Perceived health status	0.83 (0.75-0.93)*	0.94 (0.79-1.11)
Cancer fatalism	1.02 (0.99-1.05)	1.03 (0.99-1.08)
Cancer risk	1.02 (0.92-1.14)	0.97 (0.82-1.16)
Cancer fear	1.45 (1.33-1.58)*	1.28 (1.12-1.47)*
US (vs China)	0.78 (0.58-1.06)	2.03 (0.22-18.84)
Interaction terms		
US X Female	N/A[a]	0.96 (0.66-1.41)
US X Age	N/A	0.97 (0.95-0.99)*
US X High school	N/A	0.49 (0.21-1.11)
US X Vocational school	N/A	0.54 (0.24-1.18)
US X College and above	N/A	0.40 (0.18-0.90)*
US X Currently married	N/A	1.19 (0.77-1.82)
US X Currently employed	N/A	1.43 (0.95-2.14)
US X BMI	N/A	1.00 (0.96-1.05)
US X Some days	N/A	2.98 (0.87-1.16)

Variables	Model 1 odds ratio (95% CI)	Model 2 odds ratio (95% CI)
US X Not at all	N/A	1.57 (0.87-2.85)
US X Family cancer history	N/A	0.42 (0.29-0.61)*
US X $20,000 to $49,999 (50,001 to 100,000 RMB)	N/A	1.15 (0.68-1.93)
US X $50,000 to $99,999 (100,001 RMB to 150,000 RMB)	N/A	1.09 (0.59-1.99)
US X $100,000 or more (150,001 RMB and above)	N/A	1.17 (0.63-2.18)
US X Social institutions	N/A	1.03 (0.94-1.13)
US X Traditional media	N/A	1.00 (0.92-1.09)
US X Internet	N/A	0.80 (0.61-1.04)
US X Doctors	N/A	0.99 (0.79-1.25)
US X Family and friends	N/A	1.36 (1.08-1.73)*
US X Perceived health status	N/A	0.81 (0.65-1.00)*
US X Cancer fatalism	N/A	0.96 (0.90-1.02)
US X Cancer risk	N/A	1.16 (0.92-1.45)
US X Cancer fear	N/A	1.24 (1.03-1.49)*

[a] N/A: Not applicable.

*Asterisks represent the coefficients that are statistically significant at the *P*=.05 level.

Results

Descriptive Statistics

As seen in Table 2, the HINTS 2017 and HINTS-China 2017 had roughly the same percentage of cancer-free male and female respondents. However, the US sample had a much higher mean age, greater educational attainment, a lower rate of being currently married, and fewer employed people than the Chinese sample. The average BMI for the Chinese sample was lower than that of the US sample, and the percentage of regular smokers in the Chinese sample was slightly higher.

Regarding the other selected variables, the Chinese sample had higher perceived health status, while the US sample had slightly higher cancer risk perception and fear of cancer. The cancer fatalism scores for both did not differ much. The US respondents reported many more relatives diagnosed with cancers than their Chinese counterparts. The US respondents reported higher trust in social institutions, the internet, and doctors, while the Chinese respondents had more trust in family members/friends and traditional media.

Table 2. Descriptive statistics of dependent and independent variables for cancer-free respondents of the Health Information National Trends Survey (HINTS) 2017 and the HINTS-China 2017.

Variables	Categories or scales	US (N=2756)	China (N=3080)
Online cancer information seeking	Yes, n (%)	411 (14.90%)	322 (11.70%)
Predisposing variables			
Female	Yes, n (%)	1607 (58.30%)	1686 (61.17%)
Age	Years, mean (SD)	54.4 (16.1)	35.0 (11.5)
Education	Less than high school, n (%)	190 (6.90%)	494 (17.92%)
	High school, n (%)	507 (18.40%)	744 (27.01%)
	Vocational school, n (%)	813 (29.50%)	719 (26.10%)
	College and above, n (%)	1246 (45.20%)	798 (28.96%)
Currently married	Yes, n (%)	1428 (51.80%)	1944 (70.55%)
Currently employed	Yes, n (%)	1499 (54.40%)	2065 (74.94%)
BMI	kg/m^2, mean (SD)	28.45 (6.46)	22.6 (3.17)
Smoking status	Every day, n (%)	282 (10.25%)	421 (15.26%)
	Some days, n (%)	104 (3.78%)	65 (2.37%)
	Not at all, n (%)	2370 (85.98%)	2270 (82.37%)
Family cancer history	Yes, n (%)	1915 (69.47%)	620 (22.50%)
Enabling variables			
Annual family income	$0 to $19,999 / 50,000 RMB and below, n (%)	511 (18.55%)	757 (27.47%)
	$20,000 to $49,999 / 100,001 RMB to 150,000 RMB, n (%)	752 (27.27%)	1081 (39.22%)
	$50,000 to $99,999 / 100,001 RMB to 150,000 RMB, n (%)	840 (30.47%)	400 (14.52%)
	$100,000 or more / 150,001 RMB and above, n (%)	653 (23.71%)	518 (18.79%)
Trust in social institutions	3 items; 3-15, mean (SD)	8.72 (2.66)	8.49 (2.33)
Trust in traditional media	3 items; 3-15, mean (SD)	7.54 (2.62)	8.78 (2.66)
Trust in internet	1-5, mean (SD)	3.32 (1.07)	2.72 (0.75)
Trust in doctors	1-5, mean (SD)	4.51 (0.83)	3.87 (0.96)
Trust in family and friends	1-5, mean (SD)	3.10 (0.95)	3.85 (0.81)
Need variables			
Perceived health status	1-5, mean (SD)	3.41 (0.95)	3.98 (0.78)
Cancer fatalism	4 items; 4-20, mean (SD)	12.48 (3.34)	12.30 (3.19)
Cancer risk	1-5, mean (SD)	3.06 (0.96)	2.24 (0.86)
Cancer fear	1-5, mean (SD)	2.51 (1.10)	2.18 (1.01)

Logistic Regression Analysis

Table 1 presents the results of logistic regressions of the factors of online cancer information seeking for both countries. Model 1 examined the main effects, and Model 2 analyzed the moderation by country group of the associations between the three sources of predictors and internet cancer information seeking.

The results of Model 1 suggest that younger people (OR=0.99, 95% CI 0.99-1.00) who had obtained at least a college degree (OR=1.53, 95% CI 1.08-2.18), were currently not employed (OR=0.80, 95% CI 0.66-0.98), and had a family history of cancer (OR=1.51, 95% CI 1.25-1.82) were more likely to search for cancer information on the internet. In terms of enabling conditions, those who earned a moderate income (OR=0.74, 95% CI 0.58-0.94; OR=0.65, 95% CI 0.49-0.86), trusted the internet as a reliable source of information (OR=1.18, 95% CI 1.05-1.32), and distrusted family members and friends as reliable information sources (OR=0.87, 95% CI 0.78-0.97) were more likely to search the internet for cancer information. In addition, those who perceived themselves to be in poor health (OR=0.83, 95% CI to 0.75-0.93) and feared cancer (OR=1.45, 95% CI 1.33-1.58) were more likely to search the internet for cancer information.

According to the results of the tests of interaction terms in Model 2, as shown in Table 1, the Chinese and US groups differed in the associations between predisposing, enabling, and need factors and internet cancer information seeking. Three predisposing factor interaction terms were statistically significant: age (OR=0.97, 95% CI 0.95-0.99), educational attainment (OR=0.40, 95% CI 0.18-0.90) and family history of cancer (OR=0.42, 95% CI 0.29-0.61). The coefficient of family history of cancer, the only enabling factor, differed across country groups (OR=1.36, 95% CI 1.08-1.73). In terms of need factors, perceived health status (OR=0.81, 95% CI 0.65 to 1.00]) and cancer fear varied (OR=1.24, 95% CI 1.03-1.49) between the Chinese and US sample groups.

On the basis of the significance tests of the interaction terms, the researchers further calculated conditional odds ratios of predictor variables for each country group (Table 3). The relationship of age with the outcome variable differed across the country groups: senior respondents in the US sample were less likely to seek cancer information on the web (OR=0.99, 95% CI 0.98-1.00), while the opposite was true for the Chinese sample (OR=1.02, 95% CI 1.00-1.03). For the US sample, education did not play a role, while Chinese respondents with degrees from vocational schools (OR=1.58, 95% CI 1.03-2.42) or colleges and universities (OR=2.35, 95% CI 1.50-3.68) were more likely to seek cancer information online than those who had not graduated from high school. The employment status of US respondents was not a statistically significant predictor (OR=0.95, 95% CI 0.71-1.25), while a working Chinese respondent was less likely to search for cancer information online (OR=0.66, 95% CI 0.50-0.89). Chinese respondents with close and distant relatives diagnosed with cancer were more likely to search for cancer information on the internet (OR=2.23, 95% CI 1.73-2.88), which did not apply to US respondents (OR=0.94, 95% CI 0.72-1.23).

According to Table 3, the odds ratio of trust in family and friends for the Chinese sample was negative and statistically significant (OR=0.71, 95% CI 0.59- 0.86), which did not hold true for the US sample (OR=0.97, 95% CI 0.84-1.11). Although the interaction terms between family annual income/trust in the internet and the outcome variable were not statistically significant, Table 3 suggests that the conditional odds ratios for the Chinese sample were statistically significant, while those for the US sample were not. While family annual income was not related to the outcome variable for the US sample, the Chinese respondents in the two income categories between 50,001 RMB and 150,000 RMB were less likely to use the internet for cancer cognition than those below 50,000 RMB ($OR_{50,001\text{-}100,000\ RMB}$=0.70, 95% CI 0.51-0.96; $OR_{100,001\text{-}150,000\ RMB}$=0.65, 95% CI 0.43-0.99). Additionally, the conditional odds ratio of trust in the internet for the Chinese respondents was statistically significant (OR=0.71, 95% CI 0.59-0.86), while that for the US respondents was not.

The odds ratios for two need variables were statistically significant. Perceived health status was associated with internet cancer information seeking only in the US sample (OR=0.76, 95% CI 0.65-0.87]), not for the Chinese sample (OR=.94, 95% CI 0.79-1.11). Cancer fear was related to the dependent variable for both samples. However, the conditional odds ratio for the US sample (OR=1.59, 95% CI 1.41-1.79) was larger than that for the Chinese sample (OR=1.28, 95% CI 1.12-1.47), and Table 1 suggests that the difference is statistically significant.

Table 3. Bivariate and conditional odds ratios (ORs) for logistic regression of predisposing, enabling, and need variables of internet cancer information seeking.

Variable	US		China	
	Bivariate OR (95% CI)	Conditional OR (95% CI)	Bivariate OR (95% CI)	Conditional OR (95% CI)
Predisposing variables				
Female (vs male)	1.18 (0.93-1.49)	1.07 (0.83-1.37)	1.17 (0.92-1.47)	1.11 (0.83-1.49)
Age	0.99 (0.98-0.99)*	0.99 (0.98-1.00)*	1.01 (1.00-1.02)*	1.02 (1.00-1.03)*
Education (vs less than high school)				
High school	0.46 (0.23-0.90)*	0.53 (0.26-1.08)	1.01 (0.69-1.48)	1.08 (0.72-1.62)
Vocational school	0.71 (0.38-1.31)	0.85 (0.44-1.65)	1.09 (0.75-1.6)	1.58 (1.03-2.42)*
College and above	0.75 (0.41-1.35)	0.94 (0.48-1.84)	1.53 (1.06-2.2)*	2.35 (1.50-3.68)*
Currently married (vs not)	1.03 (0.82-1.30)	1.15 (0.88-1.49)	0.89 (0.70-1.13)	0.97 (0.69-1.36)
Currently employed (vs not)	1.07 (0.84-1.35)	0.95 (0.71-1.25)	0.62 (0.49-0.80)*	0.66 (0.50-0.89)*
Smoking status (vs every day)				
Some days	1.86 (0.94-3.67)	1.6 (0.78-3.28)	0.57 (0.22-1.49)	0.54 (0.20-1.46)
Not at all	1.19 (0.78-1.81)	1.48 (0.94-2.34)	1.14 (0.83-1.56)	0.94 (0.64-1.39)
Family cancer history (vs not)	1.05 (0.82-1.35)	0.94 (0.72-1.23)	2.72 (2.14-3.44)*	2.23 (1.73-2.88)*
BMI	1.00 (0.98-1.01)	0.99 (0.97-1.01)	0.98 (0.94-1.01)	0.98 (0.94-1.03)
Enabling variables				
Family annual income (vs $19,999/50,000 RMB and below)				
$20,000 to $49,999 (50,001 to 100,000 RMB)	0.78 (0.53-1.13)	0.81 (0.53-1.22)	0.79 (0.59-1.05)	0.70 (0.51-0.96)*
$50,000 to $99,999 (100,001 to 150,000 RMB)	0.66 (0.45-0.94)*	0.71 (0.46-1.09)	0.72 (0.49-1.05)	0.65 (0.43-0.99)*
$100,000 or more (150,001 RMB and above)	0.81 (0.56-1.17)	0.83 (0.52-1.34)	0.93 (0.66-1.31)	0.71 (0.48-1.06)
Trust in information sources				
Social institutions	1.02 (0.96-1.08)	1.00 (0.94-1.06)	1.01 (0.96-1.07)	.97 (0.90-1.04)
Traditional media	1.02 (0.96-1.09)	1.01 (0.95-1.07)	1.04 (0.99-1.09)	1.01 (0.95-1.08)
Internet	1.17 (0.99-1.38)	1.14 (0.99-1.31)	1.25 (1.06-1.47)*	1.43 (1.13-1.8)*
Doctors	0.97 (0.79-1.19)	0.95 (0.80-1.12)	0.94 (0.82-1.07)	0.96 (0.81-1.13)
Family and friends	0.96 (0.81-1.13)	0.97 (0.84-1.11)	0.73 (0.62-0.85)*	0.71 (0.59-0.86)*
Need variables				
Perceived health status	0.77 (0.68-0.87)*	0.76 (0.65-0.87)*	0.75 (0.65-0.87)*	0.94 (0.79-1.11)
Cancer fatalism	1.05 (1.01-1.1)*	0.99 (0.95-1.03)	1.04 (1.00-1.08)*	1.03 (0.99-1.08)
Cancer risk	1.35 (1.19-1.54)*	1.13 (0.98-1.29)	1.25 (1.09-1.43)*	0.97 (0.82-1.16)
Cancer fear	1.66 (1.49-1.85)*	1.59 (1.41-1.79)*	1.38 (1.24-1.54)*	1.28 (1.12-1.47)*

*Asterisks represent the coefficients that are statistically significant at the *P*=.05 level.

Discussion

Principal Findings

This study used HINTS 2017 of the US and HINTS-China 2017 data to compare the associations between factors related to online cancer information seeking. Under the framework of the behavioral model of health services use, the results reveal clear patterns of cross-country differences: the Chinese respondents' internet cancer information seeking was associated more with the predisposing and enabling variables, while the US respondents' information seeking was related more to the need variables. Specifically, the internet cancer information-seeking behavior of the Chinese respondents was associated with the predisposing characteristics educational attainment, employment status, and family cancer history, while that of the US respondents was not related to any of the predisposing characteristics. For enabling conditions, the internet cancer information seeking of the Chinese respondents was related to trust in family and friends as reliable health information sources, while that of the US respondents was not correlated with any enabling factor. For need variables, the internet cancer information seeking of the Chinese respondents was not related to perceived health status, while that of the US respondents was negatively associated with perceived health condition, and the magnitude of the association between cancer fear and internet cancer information seeking was stronger for the US respondents than for the Chinese respondents. These cross-country differences reveal that the extent to which predisposing characteristics, enabling conditions, and perceived needs are related to internet cancer information seeking is possibly subject to the nature of a country's cultural and structural characteristics. The importance of predisposing characteristics and enabling conditions outweighs perceived needs in countries where, for example, the culture is more collectivistic or information channel credibility is of greater concern. The perceived needs of individuals may play a larger role in more individualistic cultures.

Conventional wisdom holds that people with high socioeconomic status have high internet and health literacy, which in turn allows them to use new technologies to satisfy their cognitive and emotional needs. However, the association between being employed and information seeking suggests that working overtime and having a more active social life, by-products of being employed and financially secure in China, may limit the available time for online cancer information seeking. China has a fairly strong family-oriented culture [45]. In the context of this study, as family members remain close to each other, when certain family members are diagnosed with cancer, such news spreads very quickly within the so-called networked family. Primed by such news and vividly aware of the serious psychological and physical consequences of being a cancer patient, Chinese people might go to the internet to search for cancer-related information such as preventive measures.

Additionally, the Chinese online cancer information seekers trusted family and friends as reliable information sources more than nonseekers did, which did not apply to the US respondents. This shows that when family members are trusted as sources of information, they can displace the internet as cancer information sources in China. Although Table 1 suggests that the interaction term of trust in the internet and seeking is not statistically significant, the results in Table 3 show that the conditional odds ratio of trust in the internet was positive and statistically significant only for the Chinese sample and not for the US sample. This might be attributed to the difference in the internet between China and the US. Chinese cyberspace has long been overwhelmed with misinformation about health [51]. Thus, trust in the internet may need to become a very salient factor before Chinese people decide to use the internet as a source of cancer-related information.

The US-China differences in the associations of need factors, such as perceived personal health status and cancer fear, might reveal individualistic versus collectivistic cultural influences. Perceived personal health status was associated with online cancer information seeking only for the US respondents and not for the Chinese respondents. Previous studies have suggested that people in individualistic cultures tend to pay more attention to personal wellbeing [42]. Naturally, when perceiving their health to be deteriorating, the US respondents were more likely than the Chinese respondents to take measures such as seeking information related to cancer prevention or treatment to regain psychological or physical wellbeing. The nonsignificant results for the Chinese population are consistent with a previous study analyzing HINTS-China 2012 data [27]. In that study, the authors speculated that cancer information seeking among Chinese people might be performed only for the good of others. By resorting to a comparative design, this study can offer more convincing evidence of the close relationship between culture and health information seeking. This study also found that the odds ratio of cancer fear was much smaller for the Chinese respondents than for the US respondents, which shows that cultural characteristics might influence cancer information seeking. In individualistic cultures, negative affect is considered to be an individual's responsibility and is conceptualized as harmful, while in collectivistic cultures, negative affect is considered to be external to individuals and natural [43]. This difference sheds light on the moderating effects of culture on the correlation between negative affect (fear of cancer, in this study) and online cancer information seeking. In the context of this study, Americans' fear of cancer might result from beliefs that cancer risks are related to unhealthy personal behaviors, so they might react to cancer fears more intensely than Chinese people and strive to correct the perceived problematic situation and regain wellbeing by seeking cancer information on the internet.

In this study, differences in the associations of age and educational attainment may be related to the composition of the sample respondents. For the Americans, the higher the age, the less likely they were to search for cancer information on the web; for the Chinese, the opposite was true. As Table 2 suggests, the US sample had a much higher average age. Commonly, younger people are more technologically savvy, and older people pay more attention to personal health. In other words, age carries two types of information here. For the US sample (older age) who already paid sufficient attention to health, age was more likely to approximate internet literacy, while for the Chinese sample (younger age) who already had sufficient internet skills, age was more likely to approximate attention to health. According to Table 2, the variation in educational attainment was lower for the Americans than for the Chinese, as almost 50% of the US respondents had a college degree and above, while the educational attainment in the Chinese sample was more equally spread across the four different categories. The lack of statistically significant results may be related to the lower variance in the predictor on the American side.

Limitations

This study is not without limitations. First, as it tried to select the exact same set of variables to be compared, it excluded several important variables related to health cognition because either the Chinese or US survey did not measure them. Although the researchers made the utmost efforts to unify the measures of all the variables used in the analysis, some, such as trust in social institutions and information channels, were not identically designed, which may have impacted the results to some extent. This study is based on secondary data from HINTS. Explanations of cross-country/cultural differences involve speculation based on theoretical and practical reasoning, so future researchers are advised to further explore the precise mechanism of how culture influences internet cancer information seeking by measuring antecedents to variables such as cancer fear.

The study has other limitations as well. The researchers used two cross-sectional surveys, so causal relationships cannot be truly established. The data used in this study were only from the HINTS, and cross-country/cultural differences could be confirmed only with more replications based on additional sources of data. In this study, all the measures were self-reported, and in this context, cultural differences in response style are likely to occur, or respondents may differ in their interpretation of the questions.

Conclusions

Despite these limitations, this study makes unique theoretical and practical contributions to the literature and practice. By comparing HINTS 2017 and HINTS-China 2017 data, this study found that predisposing characteristics and enabling conditions were more associated with internet cancer information seeking for the Chinese sample and that need factors were more related to information seeking for the US respondents. Such differences might reflect possible US-China differences in job environment (eg, job pressure) and culture (individualism vs collectivism and family structure). Future health communication researchers may consider incorporating cultural values into the study design when possible. Additionally, future studies in non-Western countries may consider focusing more on predisposing factors such as structural characteristics and enabling factors related to family structure. Because international charities and health nonprofits actively promote health causes in Asian and African countries, practitioners should consider not only being culturally sensitive but also placing culture at the center of their campaigns.

Acknowledgments

This research was supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China [grant ID: 15XNQ045].

Conflicts of Interest

None declared.

References

1. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-60; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]
2. Dutta MJ. Communicating About Culture and Health: Theorizing Culture-Centered and Cultural Sensitivity Approaches. Communication Theory 2007 Aug;17(3):304-328. [doi: 10.1111/j.1468-2885.2007.00297.x]
3. Kreps GL, Yu G, Zhao X, Chou SW, Hesse B. Expanding the NCI Health Information National Trends Survey From the United States to China and Beyond: Examining the Influences of Consumer Health Information Needs and Practices on Local and Global Health. Journalism & Mass Communication Quarterly 2017 Jan 13;94(2):515-525. [doi: 10.1177/1077699016687725]
4. Health Information National Trends Survey.: National Cancer Institute; 2018. URL: https://hints.cancer.gov/data/download-data.aspx
5. Andersen R. A behavioral model of families' use of health services (Research Series No. 25). Chicago, IL: Center for Health Administration Studies, University of Chicago; 1968.
6. Andersen RM. National health surveys and the behavioral model of health services use. Medical Care 2008 Jul 8;46(7):647-653. [doi: 10.1097/MLR.0b013e31817a835d] [Medline: 18580382]
7. Song H, Omori K, Kim J, Tenzek KE, Morey Hawkins J, Lin W, et al. Trusting Social Media as a Source of Health Information: Online Surveys Comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25 [FREE Full text] [doi: 10.2196/jmir.4193] [Medline: 26976273]

8. Yang ZJ, Kahlor L, Li H. A United States-China Comparison of Risk Information–Seeking Intentions. Communication Research 2013 Mar 12;41(7):935-960. [doi: 10.1177/0093650213479795]
9. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. Journal of Health Communication 2010 Dec;15(Suppl 3):279-293. [doi: 10.1080/10810730.2010.522700] [Medline: 21154099]
10. Lagoe C, Atkin D. Health anxiety in the digital age: An exploration of psychological determinants of online health information seeking. Computers in Human Behavior 2015 Nov;52:484-491. [doi: 10.1016/j.chb.2015.06.003]
11. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. Journal of Aging and Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
12. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015 Jan 29;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
13. Oh KM, Kreps GL, Jun J, Chong E, Ramsey L. Examining the Health Information–Seeking Behaviors of Korean Americans. Journal of Health Communication 2012 Aug;17(7):779-801. [doi: 10.1080/10810730.2011.650830]
14. Nikoloudakis IA, Vandelanotte C, Rebar AL, Schoeppe S, Alley S, Duncan MJ, et al. Examining the Correlates of Online Health Information–Seeking Behavior Among Men Compared With Women. American Journal of Men's Health 2016 May 18;12(5):1358-1367. [doi: 10.1177/1557988316650625]
15. Basnyat I, Nekmat E, Jiang S, Lin J. Applying the Modified Comprehensive Model of Information Seeking to Online Health Information Seeking in the Context of India. Journal of Health Communication 2018 Jul 06;23(6):563-572. [doi: 10.1080/10810730.2018.1493058]
16. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: Changes between 2002 and 2012. Health Informatics Journal 2016 Dec 10;22(4):804-814 [FREE Full text] [doi: 10.1177/1460458215595851] [Medline: 26261218]
17. Myrick JG. The role of emotions and social cognitive variables in online health information seeking processes and effects. Computers in Human Behavior 2017 Mar;68:422-433. [doi: 10.1016/j.chb.2016.11.071]
18. Selsky C, Luta G, Noone A, Huerta EE, Mandelblatt JS. Internet access and online cancer information seeking among Latino immigrants from safety net clinics. Journal of Health Communication 2013;18(1):58-70 [FREE Full text] [doi: 10.1080/10810730.2012.688248] [Medline: 23066874]
19. Cao W, Zhang X, Xu K, Wang Y. Modeling Online Health Information-Seeking Behavior in China: The Roles of Source Characteristics, Reward Assessment, and Internet Self-Efficacy. Health Communication 2016 Feb 09;31(9):1105-1114. [doi: 10.1080/10410236.2015.1045236]
20. Magsamen-Conrad K, Dillon JM, Billotte Verhoff C, Faulkner SL. Online Health-Information Seeking Among Older Populations: Family Influences and the Role of the Medical Professional. Health Communication 2018 Feb 23;34(8):859-871. [doi: 10.1080/10410236.2018.1439265]
21. McKinley CJ, Wright PJ. Informational social support and online health information seeking: Examining the association between factors contributing to healthy eating behavior. Computers in Human Behavior 2014 Aug;37:107-116. [doi: 10.1016/j.chb.2014.04.023]
22. Yang Q, Chen Y, Wendorf Muhamad J. Social Support, Trust in Health Information, and Health Information-Seeking Behaviors (HISBs): A Study Using the 2012 Annenberg National Health Communication Survey (ANHCS). Health Communication 2016 Sep 02;32(9):1142-1150. [doi: 10.1080/10410236.2016.1214220]
23. Hayat TZ, Brainin E, Neter E. With Some Help From My Network: Supplementing eHealth Literacy With Social Ties. J Med Internet Res 2017 Mar 30;19(3):e98 [FREE Full text] [doi: 10.2196/jmir.6472] [Medline: 28360024]
24. Tustin N. The Role of Patient Satisfaction in Online Health Information Seeking. Journal of Health Communication 2010 Feb 02;15(1):3-17. [doi: 10.1080/10810730903465491]
25. Andersen RM. Revisiting the Behavioral Model and Access to Medical Care: Does it Matter? Journal of Health and Social Behavior 1995 Mar;36(1):1-10. [doi: 10.2307/2137284]
26. Andersen R, Davidson P. Improving access to care in America: Individualcontextual indicators. In: Andersen RM, Davidson PL, Kominski GF, editors. Changing the US health care system: Key issues in health services, policy, and management. San Francisco, CA: Jossey-Bass; 2001:3-30.
27. Zhao X, Mao Q, Kreps GL, Yu G, Li Y, Chou SW, et al. Cancer Information Seekers in China: A Preliminary Profile. Journal of Health Communication 2015 Apr 06;20(5):616-626. [doi: 10.1080/10810730.2015.1012244]
28. Ye Y. The Role of Illness Factors and Patient Satisfaction in Using Online Health Support Groups. Health Communication 2013 Jul 11;29(4):355-363. [doi: 10.1080/10410236.2012.757526]
29. Fei H. From the soil: The foundations of Chinese society. Berkeley, CA: University of California Press; 1992.
30. Hofstede G. Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. London: Sage; 2001.
31. Jing W, Otten H, Sullivan L, Lovell-Simons L, Granek-Catarivas M, Fritzsche K. Improving the Doctor-Patient Relationship in China: The Role of Balint Groups. International Journal of Psychiatry in Medicine 2014 Mar 17;46(4):417-427. [doi: 10.2190/pm.46.4.g]

32. Yuan B. Holistic medicine: A new medical system based on body constitution and functional status. Journal of Traditional Chinese Medical Sciences 2017 Jan;4(1):3-9. [doi: 10.1016/j.jtcms.2017.02.001]
33. Carney M. China's cancer rates exploding, more than 4 million people diagnosed in 2015, study says. , March 24. 2016. URL: http://www.abc.net.au/news/2016-03-24/chinas-cancer-rates-exploding-study-says/7272266 [accessed 2020-12-30]
34. National Cancer Center of China: Latest trend of China?s cancer in 20172018. Sohu Health. URL: http://www.sohu.com/a/220685521_792162 [accessed 2019-07-01]
35. American CS. Cancer facts & figures 2018. URL: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html [accessed 2020-12-30]
36. Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery. The Lancet 2014;384(9945):805-818. [doi: 10.1016/S0140-6736(14)61120-X]
37. Goss PE, Strasser-Weippl K, Lee-Bychkovsky BL, Fan L, Li J, Chavarri-Guerra Y, et al. Challenges to effective cancer control in China, India, and Russia. Lancet Oncology 2014 Apr;15(5):489-538. [doi: 10.1016/S1470-2045(14)70029-4] [Medline: 24731404]
38. Yu H. Reflection on the cancer treatment integrating CTM and western medicine. Journal of Traditional Chinese Medicine Management 2017;25(24):132-133.
39. Jing W, Otten H, Sullivan L, Lovell-Simons L, Granek-Catarivas M, Fritzsche K. Improving the Doctor-Patient Relationship in China: The Role of Balint Groups. International Journal of Psychiatry in Medicine 2014 Mar 17;46(4):417-427. [doi: 10.2190/pm.46.4.g]
40. Wu S, Zhu W, Li H, Lin S, Chai W, Wang X. Workplace violence and influencing factors among medical professionals in China. Am J Ind Med 2012 Nov;55(11):1000-1008. [doi: 10.1002/ajim.22097] [Medline: 22886819]
41. He AJ. The doctor–patient relationship, defensive medicine and overprescription in Chinese public hospitals: Evidence from a cross-sectional survey in Shenzhen city. Social Science & Medicine 2014 Dec;123:64-71. [doi: 10.1016/j.socscimed.2014.10.055]
42. Ahuvia A. Individualism/collectivism and cultures of happiness: A theoretical conjecture on the relationship between consumption, culture and subjective well-being at the national level. Journal of Happiness Studies 2002;3(1):23-26. [doi: 10.1023/A:1015682121103]
43. Curhan KB, Sims T, Markus HR, Kitayama S, Karasawa M, Kawakami N, et al. Just How Bad Negative Affect Is for Your Health Depends on Culture. Psychological Science 2014 Oct 10;25(12):2277-2280. [doi: 10.1177/0956797614543802]
44. Li J, Lam K, Ping Fu P. Family-oriented collectivism and its effect on firm performance: A comparison between overseas Chinese and foreign firms in china. International Journal of Organizational Analysis 2000;8(4):364-379. [doi: 10.1108/eb028923]
45. Xu A, Xie X, Liu W, Xia Y, Liu D. Chinese family strengthsresiliency. In: DeFrain J, Asay S. editors. Strong families around the world: Strengths-based research and perspectives. New York: Routledge; 2012:143-164.
46. Tang X, Wu X, Huang C, Liu R. Development Report on China's New Media. Singapor: Springer; 2017.
47. Zhang M, Zhang C, Sun Q, Cai Q, Yang H, Zhang Y. Questionnaire survey about use of an online appointment booking system in one large tertiary public hospital outpatient service center in China. BMC Medical Informatics and Decision Making 2014 Jun 9;14(1):A. [doi: 10.1186/1472-6947-14-49]
48. Amazon. com, Inc. (AMZN) lags behind Chinese tech firms in healthcare push, 2018. 2018 Jul 1. URL: https://stocknews.com/news/amzn-amazon-com-inc-amzn-lags-behind-chinese-tech-firms-in-healthcare/ [accessed 2019-12-30]
49. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. Journal of Statistical Software 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
50. Rubin D. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
51. Lu Z, Jiang Y, Lu C, Naaman M, Wigdor D. The Government's Dividend: Complex Perceptions of Social Media Misinformation in China. 2020 Apr Presented at: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; April 2020; Virtual p. 1-12 URL: https://dl.acm.org/doi/fullHtml/10.1145/3313831.3376612 [doi: 10.1145/3313831.3376612]

Abbreviations

HINTS: Health Information National Trends Survey

Edited by G Eysenbach; submitted 02.10.20; peer-reviewed by E Neter; comments to author 26.10.20; revised version received 01.12.20; accepted 19.12.20; published 11.01.21.

Please cite as:
Zhang D, Hu H, Shi Z, Li B
Perceived Needs Versus Predisposing/Enabling Characteristics in Relation to Internet Cancer Information Seeking Among the US and Chinese Public: Comparative Survey Research
J Med Internet Res 2021;23(1):e24733
URL: http://www.jmir.org/2021/1/e24733/
doi:10.2196/24733
PMID:33427668

Original Paper

A Bayesian Network–Based Browsing Model for Patients Seeking Radiology-Related Information on Hospital Websites: Development and Usability Study

Ryusuke Suzuki[1], MHS; Teppei Suzuki[1,2], MBA, PhD; Shintaro Tsuji[1], PhD; Kensuke Fujiwara[1,3], PhD, MBA; Hiroko Yamashina[1], PhD; Akira Endoh[4], PhD; Katsuhiko Ogasawara[1,4], MBA, PhD

[1]Graduate School of Health Sciences, Hokkaido University, Sapporo, Japan
[2]Iwamizawa Campus, Hokkaido University of Education, Iwamizawa, Japan
[3]Graduate School of Commerce, Otaru University of Commerce, Otaru, Japan
[4]Department of Medical Informatics, Hokkaido University Hospital, Sapporo, Japan

Corresponding Author:
Katsuhiko Ogasawara, MBA, PhD
Graduate School of Health Sciences
Hokkaido University
N12-W5, Kita-ku
Sapporo
Japan
Phone: 81 11 706 3409
Email: oga@hs.hokudai.ac.jp

Abstract

Background: An increasing number of people are visiting hospital websites to seek better services and treatments compared to the past. It is therefore important for hospitals to develop websites to meet the needs of their patients. However, few studies have investigated whether and how the current hospital websites meet the patient's needs. Above all, in radiation departments, it may be difficult for patients to obtain the desired information regarding modality and diagnosis because such information is subdivided when described on a website.

Objective: The purpose of this study is to suggest a hospital website search behavior model by analyzing the browsing behavior model using a Bayesian network from the perspective of one-to-one marketing.

Methods: First, we followed the website access log of Hokkaido University Hospital, which was collected from September 1, 2016, to August 31, 2017, and analyzed the access log using Google Analytics. Second, we specified the access records related to radiology from visitor browsing pages and keywords. Third, using these resources, we structured 3 Bayesian network models based on specific patient needs: radiotherapy, nuclear medicine examination, and radiological diagnosis. Analyzing each model, this study considered why some visitors could not reach their desired page and improvements to meet the needs of visitors seeking radiology-related information.

Results: The radiotherapy model showed that 74% (67/90) of the target visitors could reach their requested page, but only 2% (2/90) could reach the Center page where inspection information, one of their requested pages, is posted. By analyzing the behavior of the visitors, we clarified that connecting with the radiotherapy and radiological diagnosis pages is useful for increasing the proportion of patients reaching their requested page.

Conclusions: We proposed solutions for patient web-browsing accessibility based on a Bayesian network. Further analysis is necessary to verify the accuracy of the proposed model in comparison to other models. It is expected that information provided on hospital websites will be improved using this method.

(J Med Internet Res 2021;23(1):e14794) doi:10.2196/14794

KEYWORDS

web marketing; internet; hospitals; radiology; information-seeking behavior

Introduction

Hospital Web Marketing in Japan

For patients to choose a preferred hospital, information transmission using a hospital website is important because the external environment is constantly changing, owing to improvements to the internet and new competitive relationships as a result of the declining trend of patients [1,2]. It is assumed that there are 2 major reasons why information on websites has become increasingly important for hospital management. First, patients can choose a hospital based on their personal preferences. The “paternalism model,” where physicians make decisions, was previously the mainstream; however, patients have shifted to an “informed decision model,” where they make decisions themselves [3]. A previous study conducted by the Japanese Ministry of Health, Labor, and Welfare revealed that the percentage of those who responded that “I have obtained information” for the item “Is information usually obtained when you visit the hospital?” increased by 30% compared to 2011 [4]. Second, the popularization of smartphones among all age groups has led to significant lifestyle changes [5,6]. Patients can select a hospital based on the information available on a website. According to the National Federation of Health Insurance Societies [7], the information from a hospital website has the largest impact on hospital selection. The internet serves as an important source of medical information [8]. A study showed that “hospital websites” and “information websites about hospitals”’ account for a high proportion of the available credible information.

In previous research targeting hospitals in Japan, the introduction of marketing departments in hospitals has recently been proposed [9]. However, medical advertising guidelines make it difficult for hospitals to advertise as a company [10]. In today’s society, with advances in the internet, it is expected that the proportion of individuals who choose to seek information from such sources when deciding on a hospital for an examination or treatment will continue to increase, despite the strict regulations. Therefore, hospitals need their own web marketing methods.

Improve Information Provision on Hospital Websites

Through recent marketing, CRM, one-to-one marketing, and personalization aimed at increasing patient satisfaction and loyalty by reaching individual consumers through the spread of the internet and intensified competition are attracting attention [11,12]. It is therefore desirable to transmit appropriate information to the websites according to the needs of each visitor. For this reason, to improve information provisioning on websites, it is common to analyze the current state of websites using quantitative data such as access logs. As advertising regulations were adopted on websites in Japan, the study focused on whether the information posted on websites followed the regulations [13,14]. In the United States, there are weaker advertising regulations on websites than Japan; a study evaluated the accessibility of websites to provide high-fidelity information from the viewpoint of public health [15]. Few have focused on the browsing behaviors of visitors on hospital websites. Suzuki et al [16] attempted to develop a method for creating browsing behavior models and assessing the intentions of patients on hospital websites. However, no models have been proposed that allow medical staff to intuitively understand the browsing behavior of patients and propose improvements to meet the needs of visitors seeking radiology-related information. Therefore, we focused on a Bayesian network that can easily understand variable relationships.

Owing to the market trends, we decided to analyze not all visitors but specific visitors. Our research target, Hokkaido University Hospital, is one of the largest hospitals in Japan with 944 beds. Furthermore, Hokkaido University Hospital has one of the 17 proton therapy facilities in the country and therefore provides advanced medical care in the field of radiotherapy [17]. It is possible that there was a certain need at this hospital to provide information in response to citizen requests for the transmission of radiology-related information. In addition, after the Great East Japan Earthquake (March 11, 2011), several individuals have become interested in radiology-related information. Some studies have reported that few people have sufficient knowledge of radiology, and many people are concerned about exposure to radiation [18,19]. Owing to these situations, several patients have sought information on the safety and effects of radiotherapy and radiological diagnosis. Furthermore, there are various types of radiological diagnoses, such as radiography, computed tomography, and magnetic resonance imaging, which make it difficult for patients to understand each type of examination. This study conducted a behavioral analysis focusing on visitors seeking radiology-related information.

The purpose of this study is to propose a hospital website search behavior model in which medical staff can easily understand the browsing scenario by analyzing the browsing behavior of visitors who are looking for radiology-related information based on the results of a Bayesian network from the perspective of one-to-one marketing.

Methods

Scheme

A flowchart of our research is shown in Figure 1. The page that we believe the target wants to browse was defined as the “request page (RP).” In addition, a target reaching the requested page was defined as a “conversion,” and the ratio of conversion as the “conversion rate (CVR),” as shown in .We attempted to enhance the CVR using a Bayesian network with an access log. We used the website data access log of the Hokkaido University Hospital website (Figures 2 and 3), which has 208 pages. Our research focused on the patients seeking radiology-related information at Hokkaido University Hospital as the target visitors.

Figure 1. Research scheme.

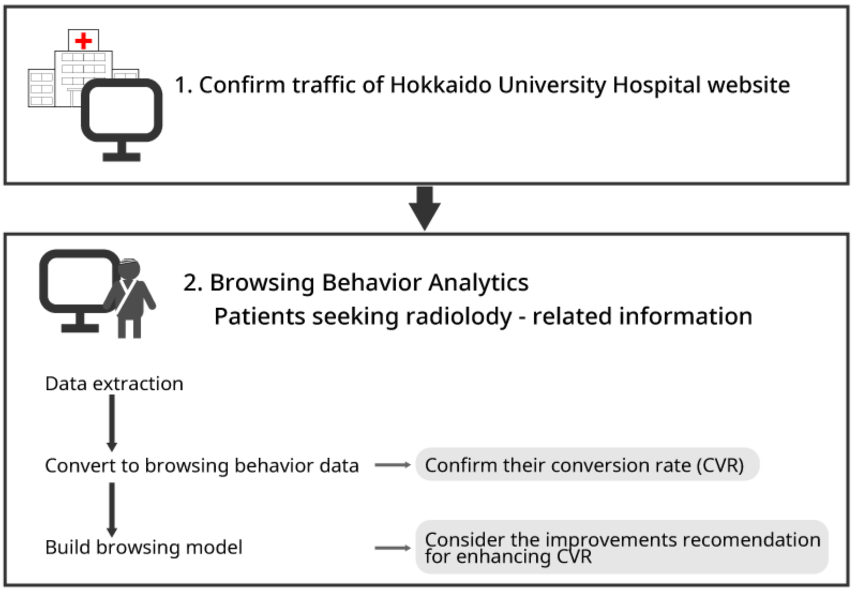


Figure 2. Abstract of target website link about radiology-related information. PRT-CT: positron emission tomography-computed tomography; RI: radio isotope.

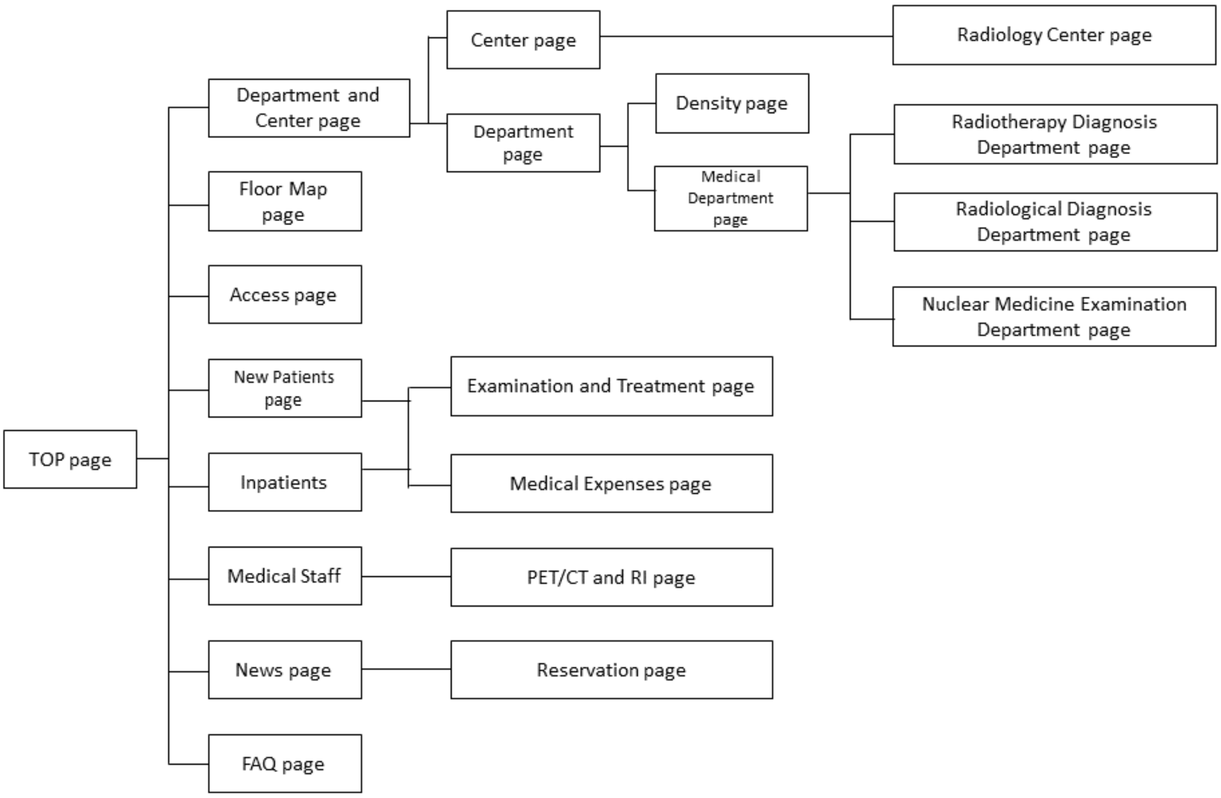


Figure 3. TOP page of Hokkaido University Hospital website (personal computer edition).

お知らせ 2015年8月18日
【重要！】平成27年10月1日(木)以降，歯科を受診する患者さんの玄関・駐車場の変更について

お知らせ 2020年2月21日

医科　受付時間

初診 予約制
午前8時30分～

再診 午前8時30分～午前11時まで

歯科　受付時間

初診 予約制
午前8時30分～正午まで

Traffic Confirmation

First, we collected the website access log of Hokkaido University Hospital from September 1, 2016, to August 31, 2017. Knowing the purpose and needs of the visitors is useful for improving a website; however, this information usually cannot be observed, and it is necessary to understand this information through an analysis of the data access log. This research analyzed the access log using Google Analytics. Google Analytics is a tool that allows users to visualize the access status without a server that even medical clerks can easily handle. Google Analytics has been widely used in public institutions, including hospitals [16,20]. We used "page view," which shows the number of pages viewed by visitors, "session," which shows the number of times any visitor accessed the website, and "page depth," which shows the number of pages viewed in a single session.

Data Extraction

First, we acquired data from the access log by designating 8 indicators using Analytics Edge, which is an add-in service of Google Analytics. The following metrics were collected from the user's http request: 1, page title; 2, retrieval keyword by visitor; 3, municipalities of visitors; 4, browsing date by visitor; 5, network domain of the visitor; 6, page depth; 7, operation system of the visitor; and 8, page views. Although Google Analytics issues a "Client ID," which identifies the user, the website administrator cannot obtain the ID. Thus, the administrator cannot identify the individual user data from the raw data. Therefore, the behavior data of each user was identified by generating a pseudo user ID from indicators 2 to 7, which led to the identification of the user. We used only indicators 3 to 7 to identify the user. We extracted data for 870 sessions using 32 radiology-related keywords that the target would use as a retrieval keyword, as shown in Multimedia Appendix 1. This selection was made by a specialist in the field of radiology. The exclusion conditions in the data extraction used were as follows: we excluded the records of the session with a page depth of 1 because of the difficulty in interpreting the intention to visit [21]. To acquire patient data, we excluded the following types of data used in retrieval keywords: recruitment-related words (eg, touring and recruitment), technical terms (eg, academic society and doctor names), and words related to needs other than treatment and examination. We also excluded data with ".ac.jp" at the end of the network domain, which is called the second-level domain.

Second, we classified the extracted sessions into 3 types of needs using retrieval keywords: 1, sought radiotherapy information; 2, sought information on nuclear medicine examination; and 3, sought information on radiological diagnosis. The judgment of the 3 classifications was based on an interpretation of indicator 1: retrieval keywords by a specialist in radiology.

Third, the RP for the 3 groups were selected, as shown in Table 1. However, the RP was an estimate because we could not know for certain what information was needed. Thus, we chose the RP in order of high scores for "Explanation of flow of examination," "Request/notes," "Explanation of technical terms," and "Explanation of equipment" in reference to the previous study, for the 3 groups from the target website [22]. Table 2 shows the evaluation criteria. The Center Navigation page, that is, the link source of the Radiology Center page, was set as RP1. Because the Radiology page was an external link, we could not examine the access log.

Table 1. Request pages in each model.

Group	Request pages			
	1	2	3	4
Radiotherapy	Center Navigation	Radiotherapy Department	—[a]	—
Nuclear medicine examination	Center Navigation	Nuclear Medicine Examination Department	Examination and Treatment	PRT-CT[b] and Nuclear Medicine
Radiological diagnosis	Center Navigation	Radiological Diagnosis Department	Examination and Treatment	—

[a] The request pages (RPs) for 3 groups were set by a specialist in radiology. The lower rank RP in a certain group was not applicable in our RP selection.

[b] PRT-CT: positron emission tomography-computed tomography.

Table 2. Criterion for page evaluation.

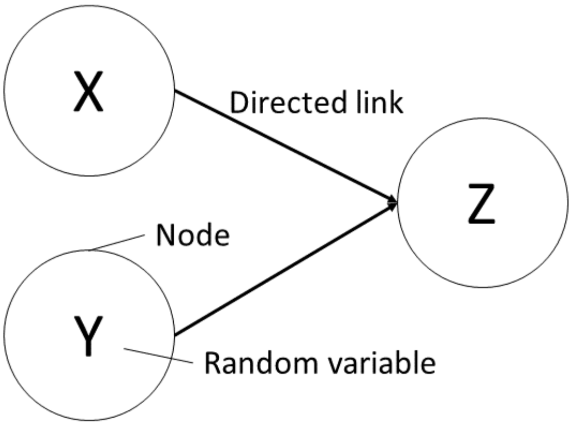
Assessment and condition	Score
Equipment description	
Statement	2
No statement	1
Doctor introduction	
Statement	2
No statement	1
Records of previous diagnoses	
Statement	2
No statement	1
Flow of examination	
Statement (image and text)	3
Statement (only text)	2
No statement	1
Precautions	
Statement	2
No statement	1
Terminology	
Statement (including meaning of the term)	2
Statement	1

Fourth, we converted the extracted access log into behavioral data. These behavioral data show whether each visitor browsed each page. After calculating CVR, these data were applied to the Bayesian network.

Browsing Models

Behavioral models were built based on the Bayesian network model, using BAYOLINK version 7.0.1. A Bayesian network is a probabilistic model that expresses the qualitative dependence among multiple random variables using a graph structure and expresses the quantitative relationship between individual variables based on their conditional probability. Variables are nodes, and a linked node is referred to as a "child node," and a link source node is referred to as a "parent node," as shown in Figure 4. In addition, Bayesian network can be used to determine the probability distributions of child node by observing the parent node. When using statistical models in marketing, it is crucial to determine whether the model is useful for decision making. Bayesian networks can express relationships between explanatory variables and even model the entire structure [23]. Although previous studies on predicting website-browsing behavior used Markov models and n-grams, such approaches were based on the premise that recent page access records have a significant influence on future page accesses [24,25]. In our models, the input value was binary for the browsing or nonbrowsing of each page for the visitors based on the premise of prior research. In other words, whether the page was browsed was represented as a node of the network. To determine the structure of the network and the parameters, we adopted Akaike's information criterion as the standard for selecting the models and the greedy search algorithm as the standard for deciding on the construction of the models. Their criterion shows adaptability to data and simplicity. The greedy search algorithm is a method of dividing the problem element into a plurality of partial problems, independently evaluating each parent combination with the highest evaluation value. In this study, all nodes were designated as parent node candidates. In addition, to show the models briefly, we built each of the models by setting pages with a small number of views as "Other pages." In our models, all the variables were binary and indicated whether browsing was performed. The joint probability indicates the probability of an arbitrary behavior being observed, and the conditional probability distribution indicates the distribution of the browsing status of the target pages considering that of a certain page. Therefore, our models show the cause-and-effect relationship between the browsing pages of the target visitor. For example, in a model in which the child node is the treatment page and the parent node is the department page, we interpret this to mean that the browsing probability for the treatment page is affected by the browsing probability for the department page. A method was then necessary to judge the strength of the influence.

Figure 4. Bayesian network outline.



After structuring the browsing model, the sensitivity analysis was performed to determine the influence of other variables in the model on any variable and searched for pages that affected the browsing probability of the RP using mutual information. Mutual information content represents a measure of the independence of 2 random variables. Referring to the results of the sensitivity analysis, hypothetical experiments called a probability inference were conducted to determine the probability distribution of any other variable when observing certain variables and predicted the resulting variable, that is, the browsing probability of the RP.

Finally, some improvements were discussed to increase the browsing probability of the RP and meet the needs indicated from the models.

Results

Traffic Summary of the Target Website

There were 2,521,279 sessions during the study period, as presented in Table 3. Access to the department pages was concentrated next to the TOP page, as listed in Table 4. After data extraction, a behavior analysis was conducted for a total of 90 sessions, including 35 sessions for the radiotherapy interest group, 35 sessions for the nuclear medicine examination interest group, and 20 sessions for the radiological diagnosis interest group. The CVR for the 3 groups is presented in Table 5. In our study, 74% (67/90) of the targets could reach the RP.

Table 3. Access during the study period.

Indication	Values
Page views, n	2,521,279
Sessions, n	663,213
Users, n	381,307
Average pages viewed per sessions	3.8
Average session time (second)	146

Table 4. Access rate of main pages and radiology-related pages.

Page title	Page views, n (%)
TOP	362,785 (14.39)
Department Navigation	187,558 (7.44)
Medical Department Navigation	153,454 (6.09)
Department and Center Navigation	123,462 (4.90)
Access	99,806 (3.96)
Center Navigation	28,190 (1.12)
Radiotherapy Department	7757 (0.31)
Radiological Diagnosis Department	4839 (0.19)
Nuclear Medicine Examination Department	4651 (0.18)
Examination and Treatment	2407 (0.10)

Table 5. Accessibility to request pages in each model.

Group	CVR[a] (%)				
	RP[b]1 (%)	RP2 (%)	RP3 (%)	RP4 (%)	Total (%)
Radiotherapy	2.9	80	—[c]	—	80
Nuclear medicine examination	2.9	31	9.0	57	83
Radiological diagnosis	0	45	5.0	—	50

[a] CVR: conversion rate.

[b] RP: request page.

[c] The lower rank RP in certain groups was not set in our judgement. Therefore, certain CVRs were treated as not applicable.

Radiotherapy Interest Group

The model described in Figure 5 was structured using 35 sessions from the learning data. The browsing of RP2, the "Radiotherapy Department page," was affected by the browsing of the Proton Therapy News page, which affected the browsing of the radiology-related department and Medical Expenses pages. The results of the sensitivity analysis are illustrated in Figure 6. The mutual information volume of the Proton Therapy News page was the largest, indicating that browsing this page had the greatest effect on the browsing of RP2. Based on the results of the sensitivity analysis, we calculated the change in the browsing probability of RP2 owing to browsing or nonbrowsing of the pages based on a probability inference, as shown in Figure 7. When the target visitor browses the Radiological Diagnosis Department page, the probability of RP2 increased by 17% from the prior probability.

Figure 5. Radiotherapy model.

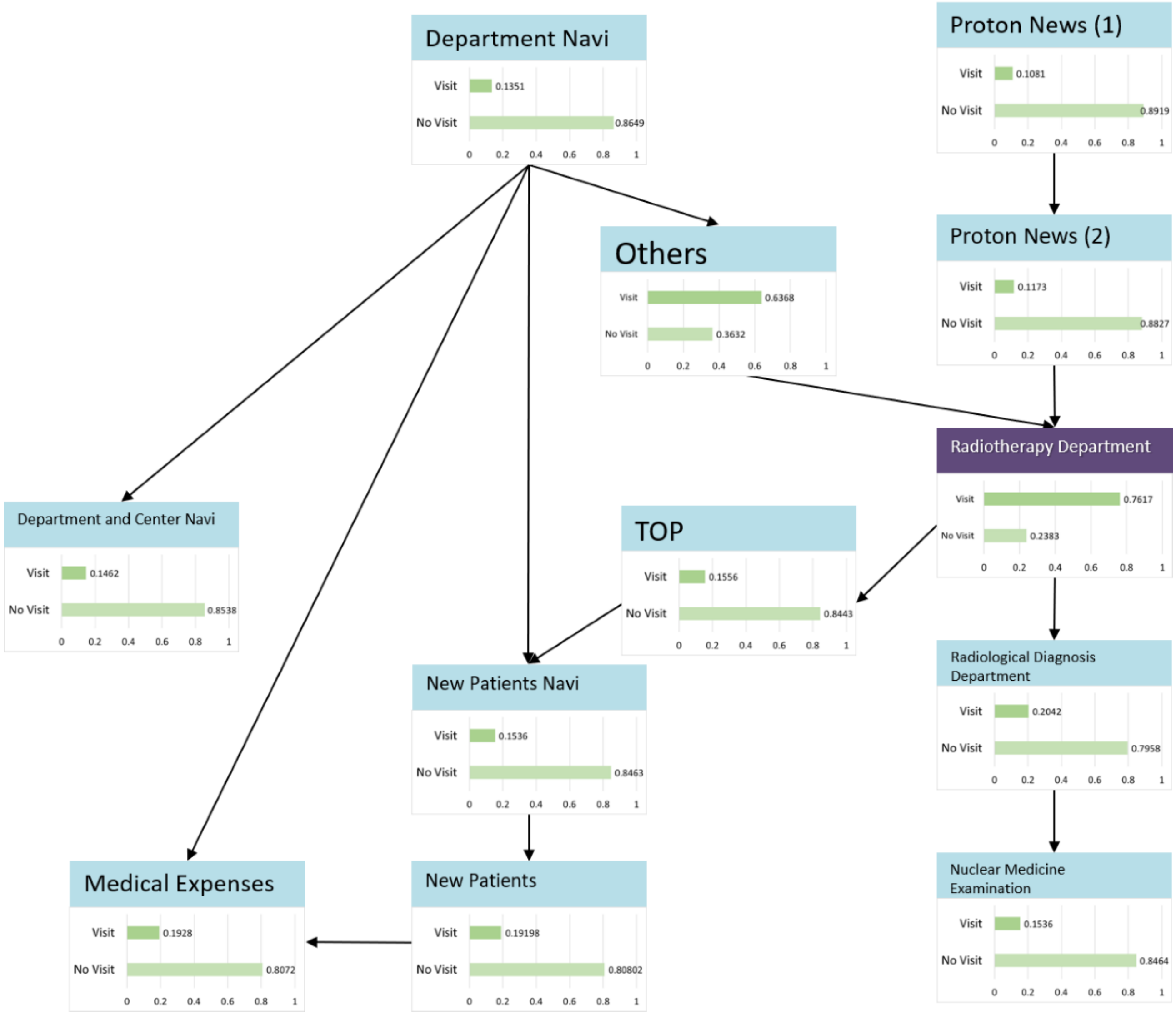


Figure 6. Sensitivity analysis (radiotherapy interest group).

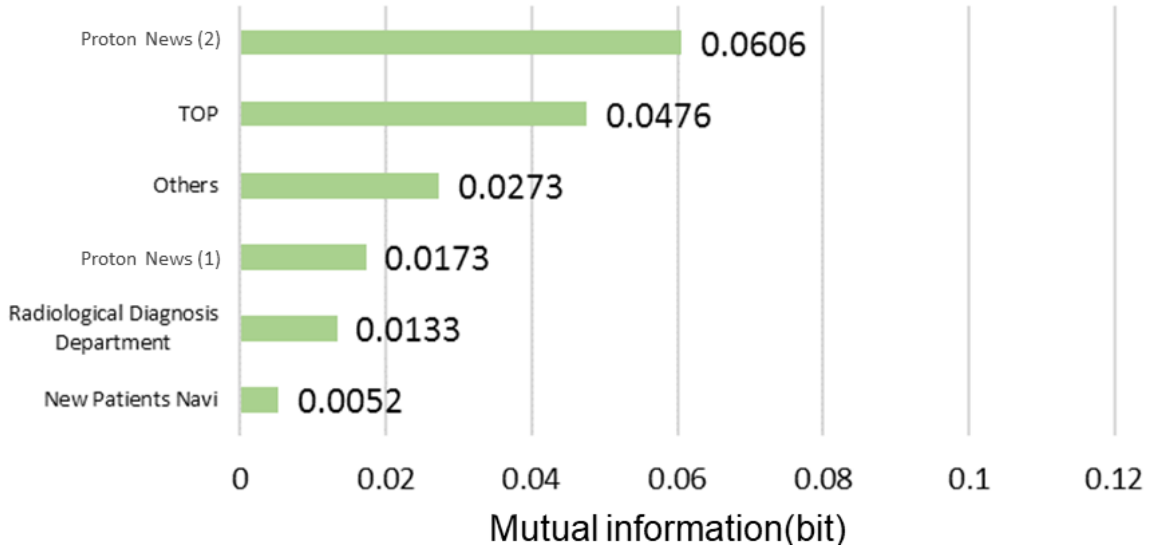
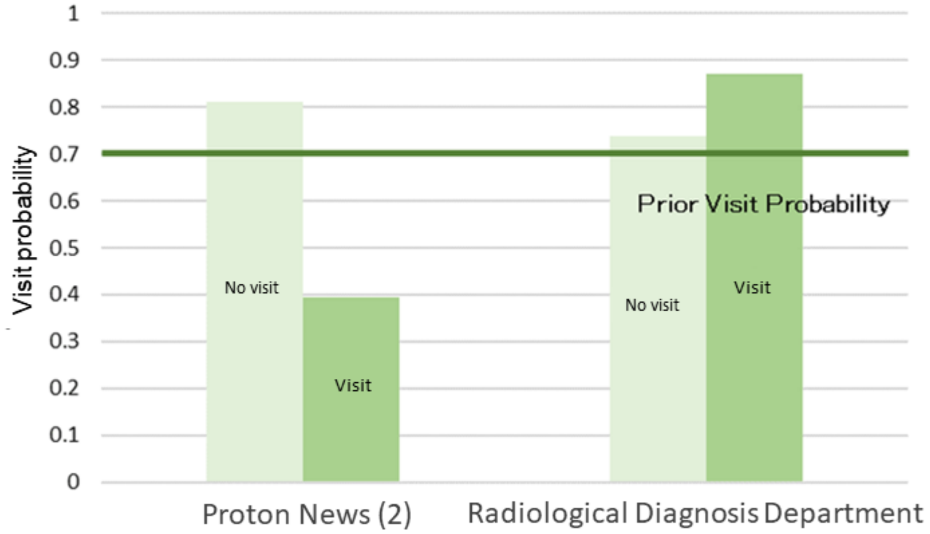


Figure 7. Probability inference (radiotherapy interest group, Radiotherapy Department page).



Nuclear Medicine Examination Interest Group

The model, as described in Figure 8, was structured using 35 sessions from the learning data. Browsing of the RP2 “Nuclear Medicine Department page” was affected by the browsing of RP4 “PRT-CT and Nuclear Medicine Examination” and was related to the browsing of the reservation and the medical staff pages. The results of the sensitivity analysis are shown in Figure 9. The interaction between the browsing of RP2 and that of RP4 was also demonstrated. We then calculated the change in the browsing probability of RP2 and RP3 owing to the browsing or nonbrowsing of the pages, as depicted in Figures 10 and 11. The browsing probability of RP2 increased when browsing the Medical Staff page. When the target visitors browsed the FAQ page, the browsing probability of RP3 increased by 15% from the prior probability.

Figure 8. Nuclear medicine examination model.

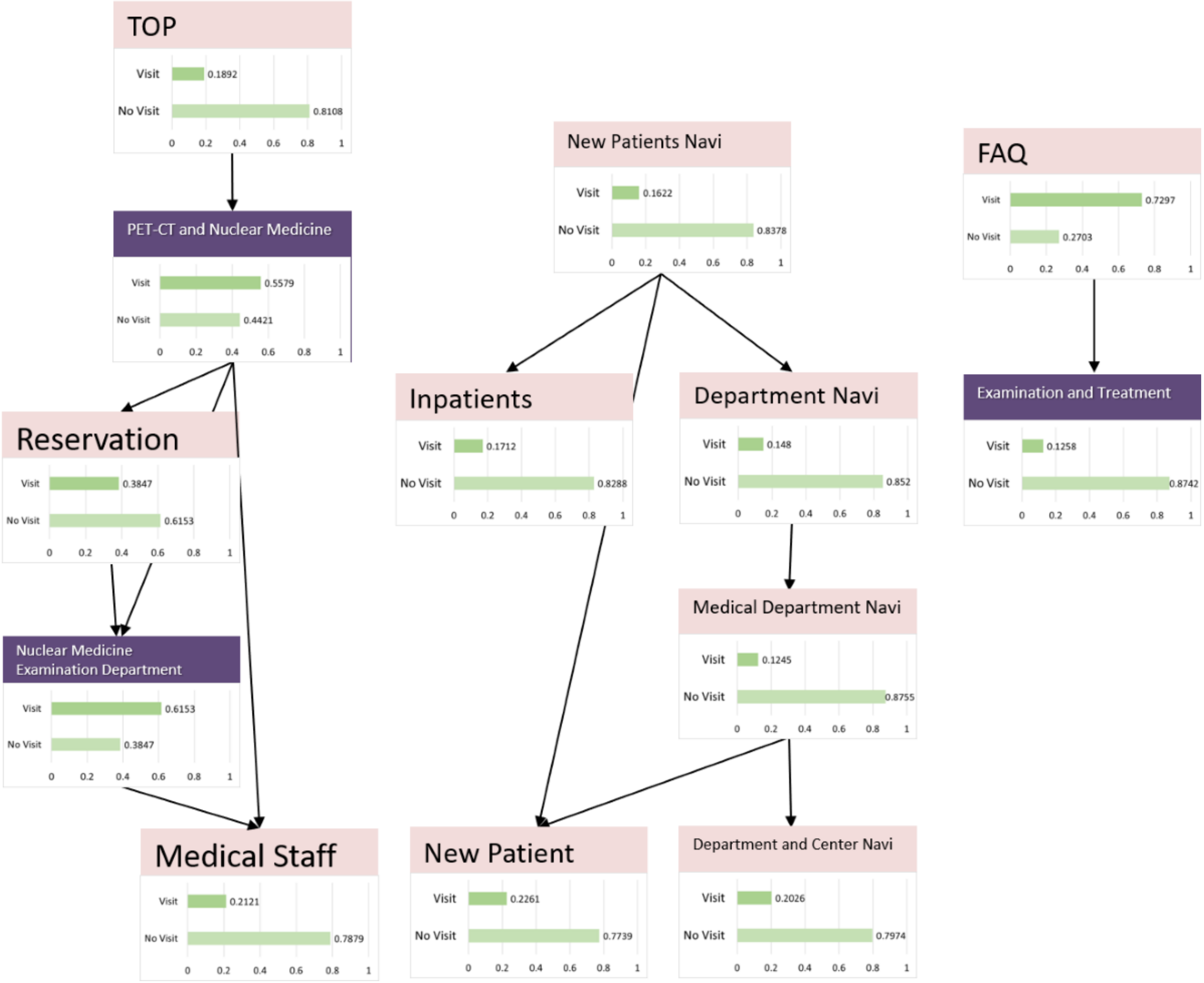


Figure 9. Sensitivity analysis (nuclear medicine examination interest group). PET-CT: positron emission tomography-computed tomography.

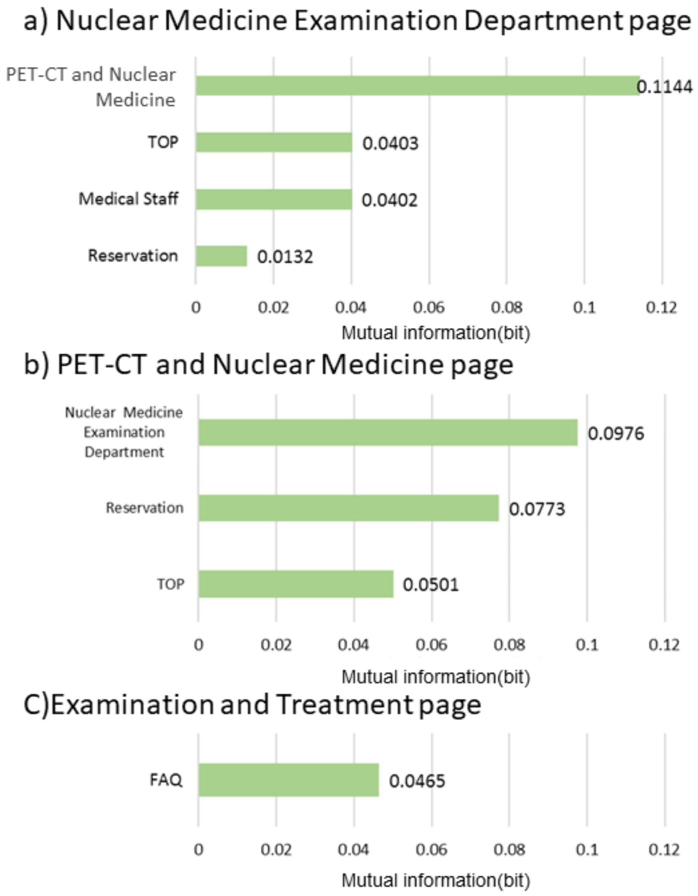


Figure 10. Probability inference (nuclear medicine examination interest group, Nuclear Medicine Examination Department page). PET-CT: positron emission tomography-computed tomography.

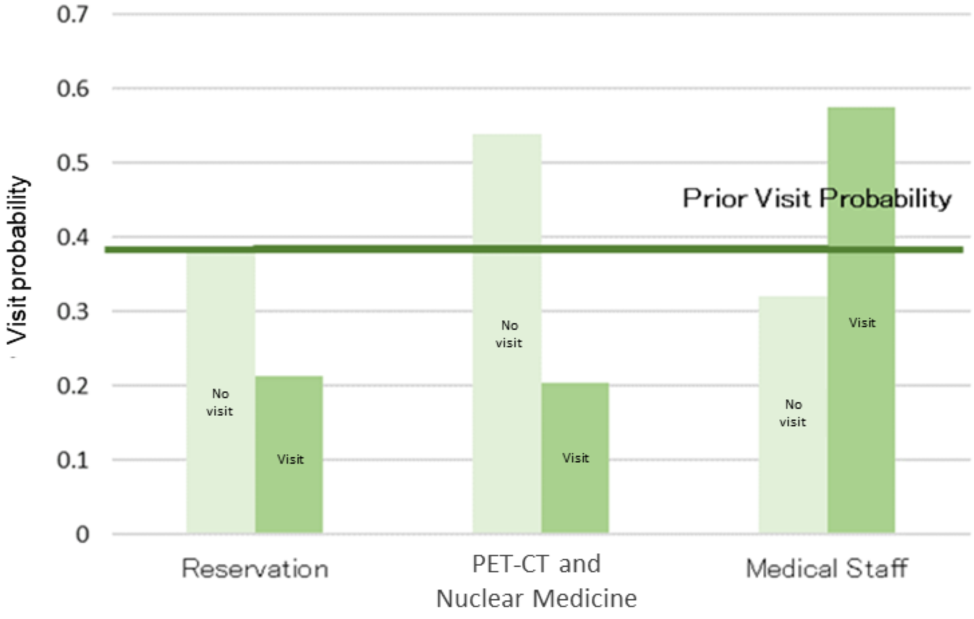
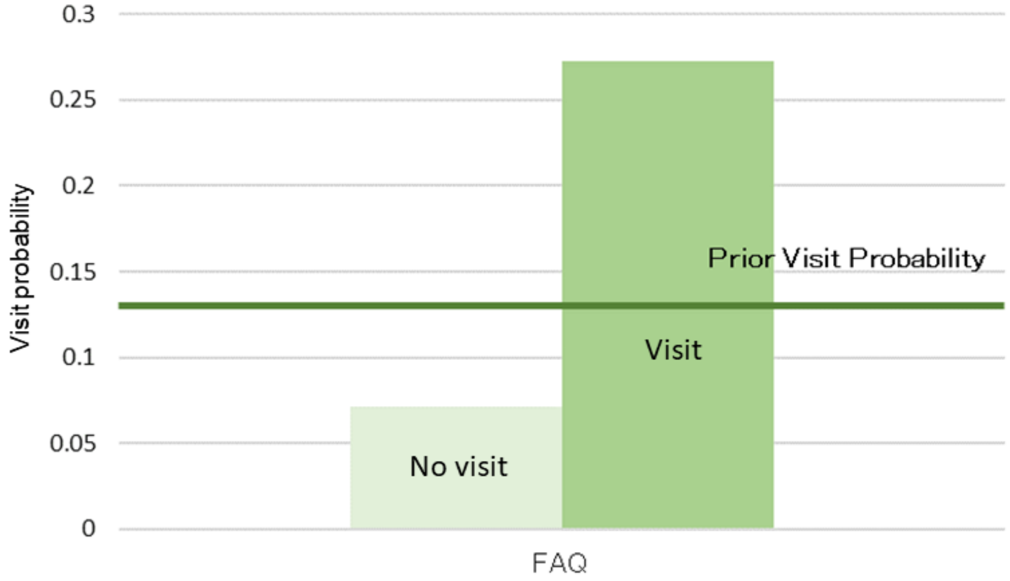


Figure 11. Probability inference (nuclear medicine examination interest group, Examination and Treatment page).



Radiological Diagnosis Interest Group

The model described in Figure 12 was structured using 20 sessions from the learning data. The browsing of the RP2 “Radiological Diagnosis Department page” affected the browsing of the Radiotherapy Department and related department pages, such as Orthopedics. The results of the sensitivity analysis are shown in Figure 13. Browsing of the Radiotherapy Department page had the greatest impact on RP2 viewing. We then calculated the change in the browsing probability of RP2 owing to the browsing or nonbrowsing of the Radiotherapy Department page, as shown in Figure 14. When the target visitors browsed the Radiotherapy Department page, the probability of RP2 increased by 35% from the prior probability.

Figure 12. Radiological diagnosis model. PET-CT: positron emission tomography-computed tomography.

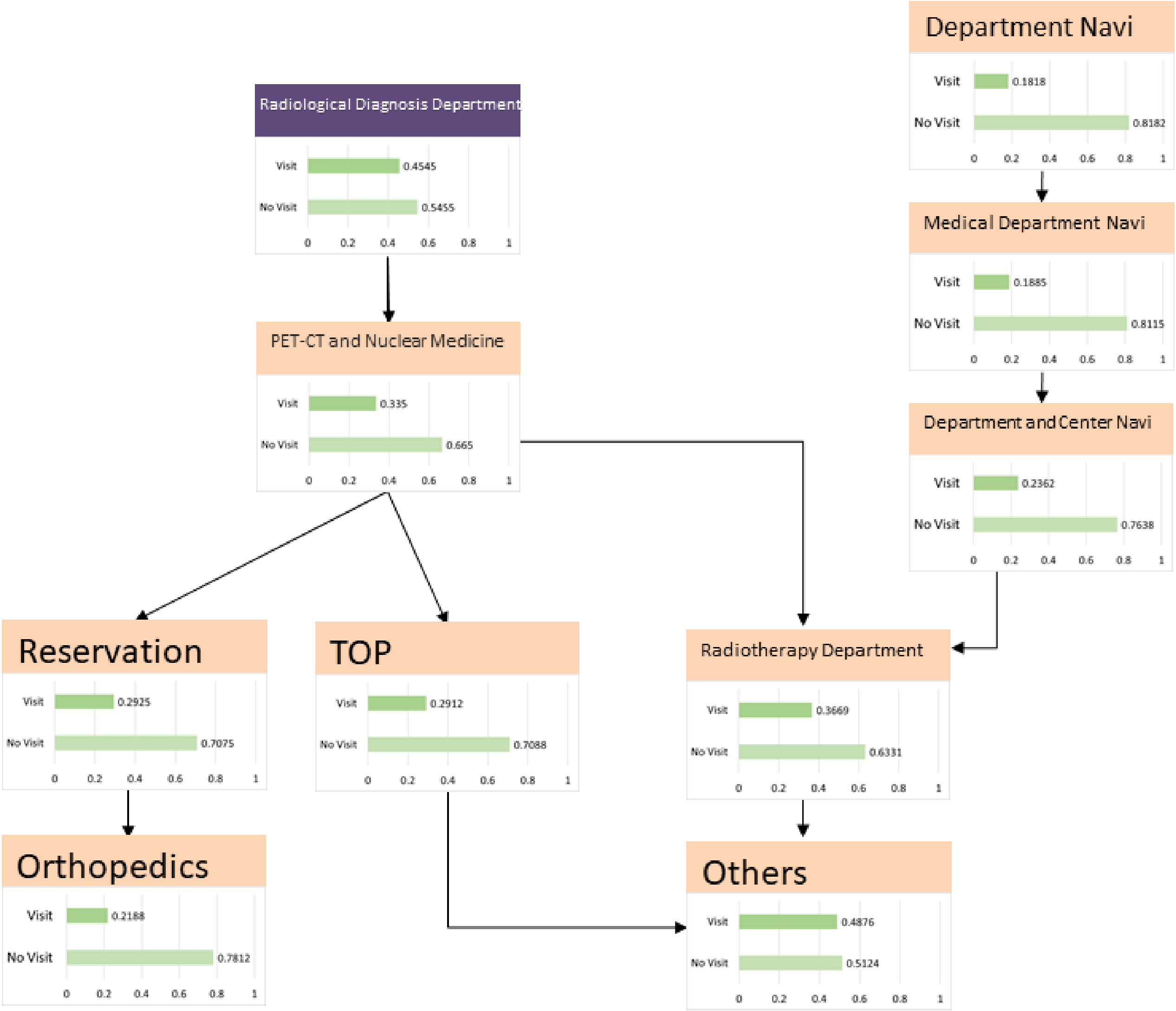


Figure 13. Sensitivity analysis (radiological diagnosis interest group).

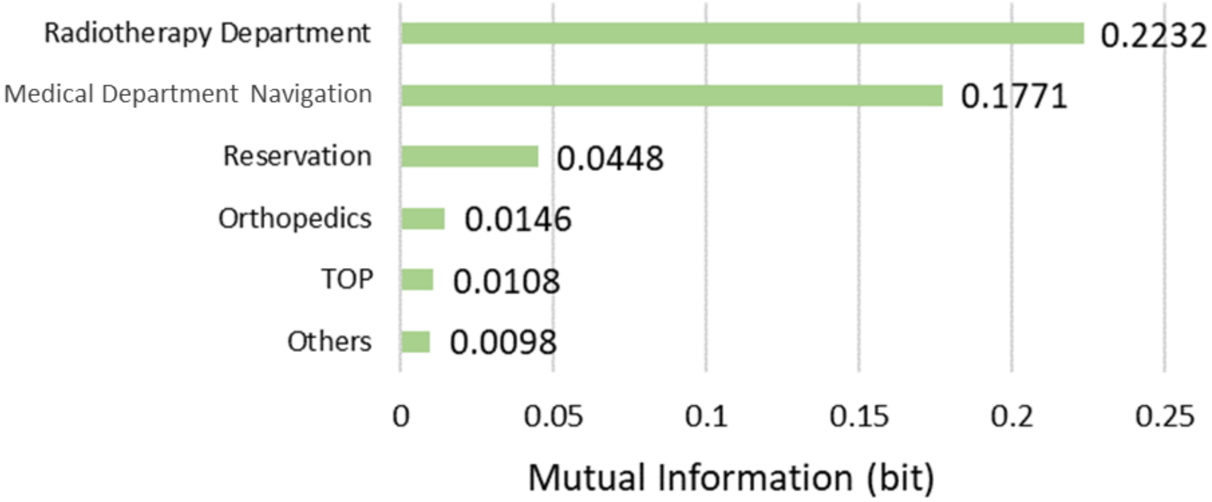
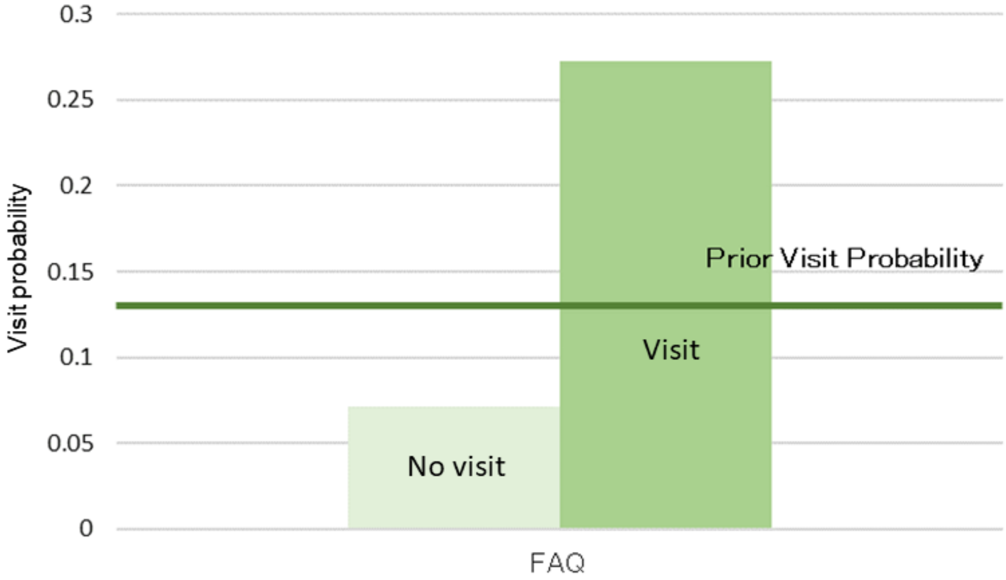


Figure 14. Probability inference (radiological diagnosis interest group, Radiological Diagnosis page).



Discussion

The number of visiting sessions on the Hokkaido University Hospital website is over 650,000 per year. This number is comparable to the number of actual patients to the hospital annually (approximately 740,000 in 2015) and indicates that the website plays a key role as an information provider for many individuals.

Only 2% (2/90) of the target visitors could reach the Radiology Center page (RP1), which describes its inspection information. There are 3 possible reasons for this finding: browsing behavior (the visitor started with the department pages, not the Center page), the visitor could not distinguish the Radiology Center pages from the radiology-related department pages that describe the department information, and that there are some pages for which we could not obtain the data access log. In total, 2 plans were considered for improving the browsing probability of RP1: linking the radiology-related department pages and the Radiology Center page, and summarizing radiology-related information on the website rather than posting according to the hospital's organizational structure.

According to the radiotherapy model, when browsing the Proton Therapy News page, the browsing probability of the RP2 "Radiotherapy Department page" decreased. Therefore, we found that interest in radiotherapy and interest in news differed from one another, or that there was no conductor design. In addition, because the child node of the Radiotherapy Department page had the pages of new patients and medical expenses pages, patients who receive radiation treatment tend to seek information on receiving medical treatment such as medical expenses owing to the high price of proton therapy [26]. In a questionnaire survey of medical radiological specialists, it was reported that a certain number of technicians were asked about the cost of radiotherapy in treating patients receiving cancer radiotherapy [27].

In the nuclear medicine examination interest group, the browsing probability of the FAQ page was higher than that of the other 2 groups. In addition, as a result of the probability inference, a positive relationship was found between browsing the FAQ page and the Examination and Treatment page. Because the FAQ page is generally used for posting information on exceptional cases such as product guidelines and specific topics [28], target visitors may be concerned about specialized information, except for the content of the RP. From Figure 10, the browsing probability of the RP2 "Nuclear Medicine Examination Department page" was higher than the prior probability when browsing the Medical Staff page. Because neither of these pages were designed with a link that allows direct access, it is mainly the medical staff that browse them.

In the radiological diagnosis model, the RP2 "Radiological Diagnosis Department page" contained child nodes, such as the Orthopedics Department page, indicating that visitors are interested in the contents of the departments they are associated with. Table 6 presents the proportion of radiological diagnosis conducted at Hokkaido University Hospital, with the proportion of "bone/soft" being the second highest after "respiratory organs." Therefore, the site design was considered, linking radiology-related department pages with their related department pages and redesigning the website considering the flow of treatment and inspection.

Table 6. Number of radiological examinations at Hokkaido University Hospital (April 2018 to March 2019).

Classification	Inspections, n (%)
Respiratory	53,322 (41.9)
Bone and soft part	30,230 (23.8)
Digestive organ	27,141 (21.3)
Tooth	16,685 (13.1)
Others	18 (0)

From the results of the probability inference, the browsing probability of the RP2 "Radiological Diagnosis Department page" was higher than the prior probability when browsing the Radiotherapy Department page. According to these results, visitors interested in radiology-related departments would access RP2, which indicates that CVR can be improved by linking radiology-related departments. The percentage of visitors who browsed the RP3 "Examination and Treatment page" was small. This is thought to be due to the low interest of the target visitor or their inability to reach the page. In the latter case, we suggest that CVR can be increased by linking to the "Radiology Department page," where the targets have a high percentage of browsing.

Table 7 presents the site design proposals to increase the CVR, as examined from the results. Thus, this research concluded that our method is effective in improving the provision of information because improving the quality of decision making is crucial in data analytics, as exemplified through business intelligence tools [29]. The proposed method can be applied to departments other than radiology-related areas.

Table 7. Proposals for site design to meet the needs of visitors seeking radiology-related information.

RP[a]	Improvement 1	Improvement 2	Improvement 3
A. Improve CVR[b] of radiotherapy interest group			
RP1: Radiology Center page	Link the Radiology Center page and radiology-related department pages	Link the Radiology Center page and pages such as the New Patients page and Re-examination page	Set tracking code
RP2: Radiotherapy Department page	Link the Radiotherapy Department page and Radiological Diagnosis Department page	Link the Radiotherapy Department page and the Medical Expenses page	—[c]
B. Improve CVR of nuclear medicine examination interest group			
RP1: Radiology Center page	Same as above	Same as above	Same as above
RP2: Nuclear Medicine Examination Department page	Link the Nuclear Medicine Examination Department page and the Medical Staff page	—	—
RP3: Examination and Treatment page	Review of posted contents on the Examination and Treatment page	—	—
RP4: PRT-CT[d] and Nuclear Medicine Examination page	Link the PRT-CT and Nuclear Medicine Examination page and the Nuclear Medicine Examination Department page	—	—
C. Improve CVR of radiological diagnosis interest group			
RP1: Radiology Center page	Same as above	Same as above	Same as above
RP2: Radiological Diagnosis Department page	Link Radiological Diagnosis Department page and the Radiotherapy Department page	Link the Radiological Diagnosis page and the Orthopedics page	—
RP3: Examination and Treatment page	Link the Radiological Diagnosis Department page and the Examination and Treatment page	—	—

[a] RP: request page.

[b] CVR: conversion rate.

[c] The certain CVRs were treated as not applicable in our behavioral analysis.

[d] PRT-CT: positron emission tomography-computed tomography.

Limitations

One of the limitations of this study is the low accuracy of the models applied. The number of samples (20-35 sessions) used to build the models was small compared with the actual usership for Hokkaido University Hospital's website. Owing to the recent shift to Secure Sockets Layer for use in search engines, the sample size has remained limited. Future studies should substitute search keywords. Furthermore, some data may have been missing because there were some pages set with the no web tracking code in Google Analytics. By increasing the sample size, factors such as browsing time can be incorporated into the model, improving the prediction accuracy for the probability distribution of each page. However, the effects of the visitor's interest on each page regarding factors such as the browsing time and browsing devices for hospital websites are still unknown. Therefore, it is necessary to clarify the effects of the browsing factors before modeling.

As another limitation, we were unable to verify the accuracy of the proposed model compared to other approaches. To evaluate our model, we need to accurately set the web tracking code and extract data without using retrieval keywords because we have obtained few retrieval data in recent years.

Third, the RP we set during the page-evaluation phase is uncertain. Thus, setting an RP objectively to obtain information by means of questionnaires to accurately understand visitor needs will be an issue in the future. Similarly, it will be possible to evaluate pages suitable for each visitor by clarifying the above relationship between visitor interests for each page and access indicators such as the browsing time. In recent years, Japan has been rapidly aging, and older adults are using hospital websites. However, previous studies have reported that age-related deterioration in visual acuity and changes in color vision affect the use of older adults to use such websites [30,31]. Therefore, we hope that the factors affecting CVR differ between older adults and young adults. In the future, we would like to consider improving the CVR by modeling the categorization of young and older adults.

Conclusions

This study structured 3 browsing models, each based on specific patient needs: radiotherapy, nuclear medicine examination, and radiological diagnosis. A total of 74% (67/90) of the target visitors could reach their requested page, but only 2% (2/90) could reach the inspection information page owing to the site structure. Furthermore, for the radiological diagnosis interest group, browsing of the "Radiological Diagnosis Department page" affected the browsing of the related department pages, such as Orthopedics.

Plans were considered for improving the browsing probability of the inspection information page such as summarizing radiology-related information on the website rather than posting according to the hospital's organizational structure. Thus, our method has the potential to increase the probability of delivering the desired information to the users. In the future, we plan to improve the accuracy of the models by using a large number of samples. We believe that an increase in the number of subjects will reveal the characteristics of the browsing scenario in more detail.

We built models with the objective of improving the provisioning of radiology-related information; however, because the structure of a hospital website is not unique, we hope to apply the proposed method to other hospitals. It is expected that the provisioning of information on hospital websites will be improved using this approach.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Radiation-related search keyword.
[PDF File (Adobe PDF File), 162 KB - jmir_v23i1e14794_app1.pdf]

References

1. The Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/toukei/saikin/hw/kanja/17/dl/01.pdf [accessed 2020-10-22]
2. Tani Y. Important factors in improving the management of small and medium-sized medical corporation hospitals. Management Studies (in Japanese) 2017;01:1. [doi: 10.24472/abjaba.88.0_F1-1]
3. Nakayama T. Evidence-based Risk and Benefit Communication for Shared Decision Making. YAKUGAKU ZASSHI 2018;138(3):331-334. [doi: 10.1248/yakushi.17-00185-5]
4. Treatment behavior survey (Japanese). The Ministry of Health, Labor and Welfare. 2017. URL: https://www.mhlw.go.jp/toukei/saikin/hw/jyuryo/17/dl/kakutei-kekka-gaiyo.pdf [accessed 2020-10-22]
5. Telecommunications usage trend survey (in Japanese). The Ministry of Internal AffairsCommunications. 2018. URL: https://www.soumu.go.jp/johotsusintokei/statistics/data/190531_1.pdf [accessed 2020-10-22]
6. Media fixed-point survey. Hakuhodo Inc. 2019. URL: https://mekanken.com/cms/wp-content/uploads/2019/05/167f92d09ab7cc5c8f1c38af84b23107.pdf [accessed 2020-10-22]
7. Survey on public opinion on medical care (in Japanese). National Federation of Health Insurance Societies. 2017. URL: https://www.kenporen.com/include/outline/pdf/chosa29_02.pdf [accessed 2020-10-22]

8. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013;14(1):15-23. [doi: 10.1177/1524839911405850]
9. Uchida T. A study from Actual Condition Survey of Public Relations Department at University Hospital in Japan. IT healthcare. 2010. URL: http://ithealthcare.jp/jnl/5.1/069.pdf [accessed 2020-10-22]
10. Medical advertisement guideline. The Ministry of Health, Labor and Welfare. 2007. URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000209841.pdf [accessed 2020-10-22]
11. Ha S, Bae S, Park S. Intelligent Marketing and Merchandising Techniques for an Internet Shopping Mall. PACIS Proceedings. 2001. URL: https://aisel.aisnet.org/pacis2001/75 [accessed 2020-10-22]
12. Montgomery AL, Smith MD. Prospects for Personalization on the Internet. Journal of Interactive Marketing 2009 May;23(2):130-137. [doi: 10.1016/j.intmar.2009.02.001]
13. Outline of the low that revises part of medical law (review of advertising regulation concerning medical care) (in Japanese). The Ministry of Health, Labor and Welfare. 2017. URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000171625.pdf [accessed 2020-10-22]
14. Fujimoto S, Komukai K, Kobu S, Sugita S, Kobayashi S. The quality of website information of private-practice hospital or clinics in rehabilitation. Japanese Physical Therapy Association, . (In Japanese) 2017;44(6):456-462. [doi: 10.15063/rigaku.11332]
15. Huerta TR, Walker DM, Ford EW. An Evaluation and Ranking of Children’s Hospital Websites in the United States. J Med Internet Res 2016 Aug 22;18(8):e228. [doi: 10.2196/jmir.5799] [Medline: 27549074]
16. Suzuki T, Tani Y, Ogasawara K. Behavioral Analysis of Visitors to a Medical Institution’s Website Using Markov Chain Monte Carlo Methods. J Med Internet Res 2016 Jul 25;18(7):e199. [doi: 10.2196/jmir.5139]
17. Kunugita N. [Investigation of the relationship between knowledge concerning radiation and the level of anxiety toward radiation in student nurses]. J UOEH 2008 Dec 01;30(4):421-429. [doi: 10.7888/juoeh.30.421] [Medline: 19086700]
18. Association for Nuclear Technology in Medicine. URL: https://www.antm.or.jp/05_treatment/04.html [accessed 2019-12-23]
19. Okazaki R, Ootsuyama A, Abe T, Kubo T. A Questionnaire Survey about Public's Image of Radiation after the Fukushima Daiichi Nuclear Power Plant Accident. J UOEH 2012;34(1):91-105. [doi: 10.7888/juoeh.34.91]
20. Nakatsu E, Tsuji Y, Tamiya S, Sudo K. Website renewalaccess log analysis related to the opening of the new hospital. The Journal of Otaru General Hospital. 2016. URL: https://barrel.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=4881&item_no=1&page_id=13&block_id=135 [accessed 2020-12-09]
21. Liu H, Keŝelj V. Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users’ future requests. Data & Knowledge Engineering 2007 May;61(2):304-330. [doi: 10.1016/j.datak.2006.06.001]
22. Suzuki T, Sasaki A, Hayashi K, Yagahara A, Ogasawara K. Medical information fulfillment degree of hospital website, Japan Institution of Medical Informatics 2015. (in Japanese) 2015;35(3):133-140. [doi: 10.14948/jami.35.133]
23. Mayomi H, Yoichi M. Marketing strategy product development using Bayesian networks;. Committee on Basic Artificial Intelligence (in Japanese). 2005. URL: http://mayomi.org/research01/haga/pdf02/BN2005haga050829.pdf [accessed 2020-12-09]
24. Chen MS, Park JS, Yu P. Efficient data mining for path traversal patterns. IEEE Trans Knowl Data Eng 1998;10(2):209-221. [doi: 10.1109/69.683753]
25. Oikonomopoulou D, Rigou M, Sirmakessis S, Tsakalidis A. Full-Coverage Web Prediction based on Web Usage Mining and Site Topology. 2004 Presented at: IEEE/WIC/ACM Int Conf Web Intell .?719; 2004; Thessaloniki Greece p. 716. [doi: 10.1109/wi.2004.10144]
26. Iwata H, Ishikawa H, Takagi M, Okimoto T, Murayama S, Akimoto T, et al. Long-term outcomes of proton therapy for prostate cancer in Japan: a multi-institutional survey of the Japanese Radiation Oncology Study Group. Cancer Med 2018 Feb 14;7(3):677-689. [doi: 10.1002/cam4.1350]
27. Mariko K. Survey on radiological technologists? interactions with patients undergoing cancer radiotherapy(in Japanese). The Journal of Japan Academy of Health Sciences. 2015. URL: https://www.health-sciences.jp/journal/backnumber/017_04.pdf#page=19 [accessed 2020-12-09]
28. Romero M, Moreo A, Castro J. A cloud of FAQ: A highly-precise FAQ retrieval system for the Web 2.0. Knowledge-Based Systems 2013 Sep;49:81-96. [doi: 10.1016/j.knosys.2013.04.019]
29. Wang C. A novel approach to conduct the importance-satisfaction analysis for acquiring typical user groups in business-intelligence systems. Computers in Human Behavior 2016 Jan;54:673-681. [doi: 10.1016/j.chb.2015.08.014]
30. Saito D, Saito K, Notomi K, Saito M. Comparison of standard link color visibility between young adults and elderly adults. Electron Comm Jpn 2008 Jul;91(7):41-47. [doi: 10.1002/ecj.10126]
31. Sri KZ. Research-derived web design guidelines for older people. 2005 Presented at: the 7th international ACM SIGACCESS conference on Computersaccessibility (Assets '05). Association for Computing Machinery, New York, NY, USA. ;129?135; 2005; New York, NY, USA. [doi: 10.1145/1090785.1090810]

Abbreviations

CVR: conversion rate
RP: request page

Edited by G Eysenbach; submitted 24.05.19; peer-reviewed by J Ma, IV George, E Neter, D Chao; comments to author 03.10.19; revised version received 28.02.20; accepted 18.11.20; published 19.01.21.

Please cite as:
Suzuki R, Suzuki T, Tsuji S, Fujiwara K, Yamashina H, Endoh A, Ogasawara K
A Bayesian Network–Based Browsing Model for Patients Seeking Radiology-Related Information on Hospital Websites: Development and Usability Study
J Med Internet Res 2021;23(1):e14794
URL: https://www.jmir.org/2021/1/e14794
doi:10.2196/14794
PMID:33464211

Original Paper

Effects of Objective and Subjective Health Literacy on Patients' Accurate Judgment of Health Information and Decision-Making Ability: Survey Study

Peter Johannes Schulz[1], PhD; Annalisa Pessina[1], MSc; Uwe Hartung[2], PhD; Serena Petrocchi[1], PhD

[1] Institute of Communication and Health, Università della Svizzera italiana, Lugano, Switzerland

[2] Università della Svizzera italiana, Lugano, Switzerland

Corresponding Author:
Peter Johannes Schulz, PhD
Institute of Communication and Health
Università della Svizzera italiana
Via Guiseppe Buffi 13
Lugano, 6900
Switzerland
Phone: 41 58 666 4724
Email: schulzp@usi.ch

Abstract

Background: Interpreting health information and acquiring health knowledge have become more important with the accumulation of scientific medical knowledge and ideals of patient autonomy. Health literacy and its tremendous success as a concept can be considered an admission that not all is well in the distribution of health knowledge. The internet makes health information much more easily accessible than ever, but it introduces its own problems, of which health disinformation is a major one.

Objective: The objective of this study was to determine whether objective and subjective health literacy are independent concepts and to test which of the two was associated more strongly with accurate judgments of the quality of a medical website and with behavioral intentions beneficial to health.

Methods: A survey on depression and its treatments was conducted online (n=362). The Newest Vital Sign was employed to measure objective, performance-based health literacy, and the eHealth Literacy Scale was used to measure subjective, perception-based health literacy. Correlations, comparisons of means, linear and binary logistic regression, and mediation models were used to determine the associations.

Results: Objective and subjective health literacy were weakly associated with one another (r=0.06, P=.24). High objective health literacy levels were associated with an inclination to behave in ways that are beneficial to one's own or others' health (Exp[B]=2.068, P=.004) and an ability to recognize low-quality online sources of health information (β=–.4698, P=.005). The recognition also improved participants' choice of treatment (β=–.3345, P<.001). Objective health literacy helped people to recognize misinformation on health websites and improved their judgment on their treatment for depression.

Conclusions: Self-reported, perception-based health literacy should be treated as a separate concept from objective, performance-based health literacy. Only objective health literacy appears to have the potential to prevent people from becoming victims of health disinformation.

(J Med Internet Res 2021;23(1):e20457) doi:10.2196/20457

KEYWORDS

health literacy; Newest Vital Sign; eHealth Literacy Scale; self-reported health literacy; perception-based health literacy; objective health literacy; performance-based; depression; mental health

Introduction

Background

Patients' knowledge of health matters and their capacity to acquire such knowledge have recently become more important than ever [1]. This has to do with the immense accumulation of health knowledge in medical science and, consequently, the medical profession since the enlightenment. More recently than that, a new ideal emerged, which was to replace older paternalistic conceptions of the relationship between physicians and their patients [2,3]. In the earlier concept, well-meaning physicians based their decisions on their far superior medical knowledge, experience, physical examination, and understanding of what their patients conveyed about their symptoms [3]. Based on this information, physicians developed a diagnosis and prescribed medication accordingly. In the new ideal of patient autonomy, physicians make suggestions to patients, and patients choose which suggestions to follow [3]. Other autonomy concepts include the expert patient, shared decision making, and patient-centered health care [4].

Patient autonomy requires that patients are knowledgeable, which is another reason why knowledge and knowledge acquisition in the field of health are so important today. In addition to knowledge, functional reading and writing abilities, adequate information processing, critical thinking skills, and the capability to make decisions that are beneficial to one's health [5] are all elements of health literacy. Health literacy is an immensely successful concept in health communication and public health [6]. Its success may be due in part to an intuitively felt paradox that it would be absurd indeed to train health care providers for years and produce expensive pharmaceuticals and high-tech machinery only to witness that patient outcomes are not as good as they could be because patients do not understand what their physician says and physicians fail to notice patients' lack of understanding [4]. In a sense, the existence of the term "health literacy" is the admission that not all is well with the distribution of medical knowledge in contemporary societies. Not everyone is knowledgeable enough to meet the requirements of the ideal of patient autonomy or capable enough to participate in the communication processes demanded by health care today [7].

When health literacy was first conceptualized, nobody could have known how digital communication would soon change health communication. Due to its extraordinary potential of overcoming traditional limitations to information finding, the internet is today a primary focus of health literacy research [8]. In some respects, it provides solutions or answers to problems associated with the distribution of health knowledge and the spread of health information.

In spite of the many opportunities the internet offers in providing information, social support, conversation exchange, and more advice on prevention than one could ever use, it has also magnified the quality issue in medical care for laypersons who have contact with it. The quality of the information found online is sometimes deplorable [9], and that creates its own problems. For individuals and organizations, advocating the potential benefits offered by information technology is fairly frustrating. A new, two-sided medium was introduced, which offered a way to seek and find health information more or less everywhere, at any time, on every subject, and at a very low cost. However, what people have put on the medium is, to a considerable degree, medical disinformation.

Fortunately, a solution was at hand: health literacy. If health literacy were to include the ability to discriminate true and reliable health information websites from fraudulent and misleading ones, information seekers online could learn which sources provide high- or low-quality information and could base their conclusions about the quality, and hence the credibility, of specific information on the general quality of the source. Individuals who did not have adequate health literacy could receive training to increase their health literacy or physicians could talk to them on the patient's level. This raises the question of whether individuals with a high level of health literacy also have the capability of appraising the general quality of digital sources of health information.

The hope is that when people with adequate online health literacy encounter low-quality health information, they will recognize it for what it is, and when a website is of low quality and the patient knows it, they will not accept its content and not change their stored knowledge of the subject. Some might even scan their existing knowledge when they meet a website with health disinformation and erase the knowledge that is associated with the problematic website newly encountered. That is to say, people with high health literacy might indeed be protected against the influence of disinformation and, in favorable conditions, even improve their knowledge. In contrast, people with inadequate eHealth literacy will not recognize the low quality, accept its content, and store it as newly gained knowledge of the subject.

Although people's awareness of their health literacy has not yet been considered, it should be. People might perceive themselves as more health literate than they actually are, and they might act on that misperception. It is obvious that an exaggerated misperception of one's own health literacy cannot be a good advisor for medical judgment, decisions, and behaviors. Earlier publications referred to these persons as "dangerous self-managers" [10]. Misperception of one's self is a rather well-known phenomenon in social psychology, where it exists under labels such as optimistic bias, third-person perception, and looking-glass perception.

Although health literacy is conceptualized as an objective feature, it has increasingly been measured subjectively—that is, as a self-perception—in academic research [11]. Measures of health literacy based on self-perceptions are often problematic because people might deceive themselves about their health literacy abilities [12]. At the same time, such measures gain ground in research. One example is the Brief Health Literacy Screen (BHLS), which asks the respondent 3 questions about their confidence in understanding medical information when they seek help from health care institutions [13]. Another example is the eHealth Literacy Scale (eHEALS), which consists of 8 items measuring consumers' knowledge, comfort, and perceived skills as they relate to finding, evaluating, and applying electronic health information to health problems [14].

The final example is the European Health Literacy Survey Questionnaire (HLS-EU-Q47), which assesses how easy or difficult the patient perceives health-related tasks to be [15].

Most commonly, the objective, performance-based measures require individuals to demonstrate their knowledge and familiarity regarding medical terms (eg, Rapid Estimate of Adult Literacy in Medicine [REALM] [16]) or their understanding of sentences by their ability to fill in gaps within them (eg, Test of Functional Health Literacy in Adults [TOFHLA] [17]), or to complete tasks demanding numerical ability (TOFHLA [17], STAT-interest and STAT-confidence scales [18], and Lipkus Numeracy Scale [19]). Self-reported measures of health literacy involve individuals describing themselves and their skills. Self-reported measures target the individuals' self-perceived ability to find information and understand it, and to assess individuals' confidence in their own health literacy. Subjective measures of objective concepts have long been used in survey research, often for concepts that could be openly labeled, such as knowledge or intelligence [20]. Subjective impressions of how much of these qualities individuals think they have can be ascertained directly because the concepts have spilled over into normal, nonacademic language. That has not yet happened to the concept of health literacy. Inquiry into individuals' subjective perception of their health literacy must, therefore, be performed indirectly and include questions about behaviors and processes that indicate health literacy. Self-reported measures are much more susceptible to individual and cultural influences, such as social desirability and beliefs about health and illness [21]. Conversely, a performance-based measure assumes that it is unlikely that someone would purposefully cheat on it by intentionally appearing less capable than they are [21].

The validity of a self-perceived indirect measure of health literacy can, therefore, be judged by comparing its results with the results of performance-based instruments.

Self-report and performance-based measures of health literacy have been extensively investigated in their ability to predict a range of health-related behaviors. One study [22], for example, demonstrated that performance-based numeracy and literacy, considered together, predicted skills to perform health tasks among older adults. Along the same lines, Gazmararian et al [23] showed that patients with inadequate performance-based health literacy were less likely to adhere to their physicians' treatment a year later. People with limited health literacy were less likely to have sought cancer information and more frequently endorsed fatalistic beliefs about their decision making [24].

On the other hand, even low levels of self-reported health literacy seem to be associated with worse health outcomes. One study in patients with diabetes [25] found that low self-reported health literacy was associated with less diabetes knowledge, glycemic control, and physical activity. Mitsutake et al [26] found mixed results. Individuals with high self-reported health literacy were significantly more likely to exhibit good health behaviors, perform physical exercise, and eat a balanced diet. Still, there were no associations of health literacy with cigarette smoking, alcohol consumption, hours of sleep, or eating between meals. Wångdahl et al [27] found that refugees with limited self-reported health literacy reported worse health, impaired well-being, and fewer health care–seeking behaviors than those with high health literacy.

A systematic review [11] found a paucity of research analyzing the relationship between objective and subjective health literacy and health outcomes. Haun et al [28] found no significant relationship between limited health literacy and patient self-report of having or not having hypertension, diabetes, or a past stroke after adjusting for covariates (ie, gender, race, education, self-reported reading level, retiree status, and having a functional disability). Hirsh et al [29] found a significant association between the self-reported measure of health literacy and a multidimensional measure of health, which did not hold with a performance-based tool. Kiechle et al [11] found no differences in the relationship between performance-based and self-reported health literacy for 4 of 6 outcomes (self-reported diabetes, stroke, hypertension, and a physician-completed rheumatoid arthritis disease activity score).

Hypotheses

The broader question that this research addressed was the utility of self-perceived health literacy in health communication research. The considerations outlined above suggest three comparisons that should help to assess this utility: (1) the association of self-perceived and objective health literacy, (2) the influence of health literacy on choice of behavior, and (3) the ability to recognize disinformation on the web. These variables were seldom studied together and the results were inconsistent [30-36]. In a unique study by Benotsch et al [37], they found an overall positive association between low health literacy levels and wrong perceptions of information quality. Specifically, low health literacy was associated with lower quality ratings of a high-quality website and higher quality ratings of a low-quality website.

Therefore, our primary research question (RQ1) was as follows:

- RQ1: How strongly will a self-perceived measure of health literacy be associated with a performance-based measure?

Four hypotheses were formulated to present the associations in a testable format:

- H1: Participants who score high on an objective, performance-based test of health literacy will make better treatment choices for depression than participants who score low.
- H2: Participants who score high on an objective, performance-based test of health literacy will make better judgments on the quality of mental health websites.
- H3: Participants who score high on an objective, performance-based test of health literacy will make better treatment choices based on their judgments of the quality of mental health websites, reflecting a mediation of the relationship between health literacy and depression-related decision making by the perception of website quality.
- H4: Self-reported health literacy will not produce similar associations.

Methods

Study Design

The study is part of a larger experiment testing an aspect of the order of persuasive communication content. This experiment and the order of presentation are of no concern to this paper. Therefore, we did not include order as a variable here.

The experimental design, however, had consequences for the structure of the data collection. We considered it important for the experiment to have roughly equally sized groups of high and low objective health literacy. For this purpose, data were collected in two steps. The first step involved an assessment of objective health literacy and the collection of sociodemographic information. The findings of the first step were used to define high and low literacy and to assign participants into experimental groups randomly. The second step consisted of all other measures. Two pilot studies were carried out before the study was conducted (see Multimedia Appendix 1).

Variables and Measures

The independent variables were two measures of health literacy: (1) a subjective, perception-based measure, and (2) an objective, performance-based measure. The measures were available as total scores or as dichotomous summaries (low/inadequate versus high/adequate level). An intervening variable was the perception of the quality of two websites, one of dubious quality, the other of high quality. The dependent variable was treatment choice, considered as a dichotomous variable. Several covariates were also considered.

Performance-based health literacy was measured using the Newest Vital Sign (NVS) [38], which uses a fictitious ice cream label. It measures literacy, comprehension, numeracy, application/function, and evaluation skills by asking questions that are answered after a person observes the label. The final score ranges from 0 (limited literacy) to 6 (adequate literacy); the dichotomy separated participants with low/inadequate health literacy (scores of 0-4) from those with high/adequate health literacy (scores of 5-6). The scale demonstrated good internal reliability (α=.62).

Self-reported health literacy was measured using the eHealth Literacy Scale (eHEALS) [14,39]. Online health literacy, or eHealth literacy, refers to “the ability to seek out, find, evaluate and appraise, integrate, and apply what is gained in electronic environments toward solving a health problem” [40]. The scale is an 8-item, self-reported measure developed to assess consumers’ combined knowledge, comfort, and perceived skills at finding, evaluating, and applying eHealth information to health problems. Responses are measured using 5-point Likert scales whose options range from 1 (“strongly disagree”) to 5 (“strongly agree”). The total scores were calculated as the mean, with the ratings formatted such that higher scores (ie, closer to 5) represented higher self-reported eHealth literacy. The scale demonstrated good internal reliability (α=.89) and internal consistency (r_s>0.58; mean 3.17, SD 0.78). The dichotomous form was computed based on the median split (median 3.25), thus generating one group with low (n=178) and one group with high (n=184) self-reported health literacy.

Website Quality

Two versions of a mock website informing about treatment options for depression were produced, one of high quality and the other of low quality. To provide our study with ecological validity, we retrieved the contents from real websites, one of high quality [41] and the other of low quality [42]. Both appeared among the top results on Google.it when the keywords “anti-depressive” and “natural remedies for depression” were searched. The criteria that we used to identify a high-quality website were having a main focus on the most effective treatment options according to scientific research (namely antidepressant medication and psychotherapy), and a clear debiasing intention toward the most popular misconceptions about them [43]. Conversely, the criterion for low-quality was the emphasis on common misbeliefs about treatment [43], including the overestimation of the curative effect of natural and self-help remedies [44]. However, in contrast with the real websites, the amount of information on the mock websites was shortened in order to make the stimuli comparable in terms of number and pieces of information provided, number of words, and overall cognitive effort required. Moreover, we used the same layout for both stimuli (eg, color, graphics) and a simple screenshot instead of an interactive website in order to limit the effects due to uncontrolled variables (see Multimedia Appendix 2). A 1-item manipulation check was included to identify if people had read, understood, and could recall the information on the website.

Perceived Information Quality

Participants’ perceptions of information quality were assessed using 7-step semantic differential scales [30,45]. Positive adjectives were “accurate,” “reliable,” “complete,” and “understandable,” while the respective opposites were “inaccurate,” “unreliable,” “incomplete,” and “nonunderstandable.” High numbers on the scales indicated positive perceptions. The data showed good internal reliability (α=.88 for high-quality websites, and α=.90 for low-quality websites) and internal consistency (r_s>0.68 for high-quality websites, and r_s>0.75 after removing the item understandability). Consequently, the scores for reliability, accuracy, and completeness were considered in the analyses.

Outcome Variable

From a list, participants were asked to choose one or more treatments for depression for themselves (help-seeking behavior), a family member, and a close friend (advice-giving behavior). The Kuder-Richardson 20 coefficient indicated good internal consistency among the three items (KR20=0.96), thus providing the basis to develop a single measure of the construct. The final score ranged from 1 (correct choice) to 0 (wrong choice). Examples of the right choices were psychotherapy, antidepressant medications, and seeking help from a doctor. Examples of wrong decisions were treating depression with St John’s wort or vitamins or yoga without any mention of antidepressant medications or psychotherapy.

Covariates

The frequency of online health-related information–seeking behavior in general as well as about depression in particular were both asked with a single item. The response options ranged from 1=never to 6=more than two times a week. During the scoring procedure, both items were reversed, with higher scores indicating lower frequency. Participants' previous experience with depression was assessed in terms of having ever suffered from depression or helping someone close to them, such as a family member or a close friend, who was suffering from depression. The response options were "no previous experience" or "yes previous experience."

Sample

The study was conducted in Italy using a snowball sampling method by posting the link to the survey on several public and private Facebook pages. The survey was implemented on QualtricsXM software (version 2019; Qualtrics). The inclusion criteria for study participation were adults aged between 18 and 65 years who were residing in Italy, with a good command of the Italian language and internet access. Mental health workers (psychiatrists, psychologists, and psychotherapists) and psychology graduates/students were excluded from the sample.

An a priori power analysis was conducted using G*Power 3.1.9.4 software [46] to determine the sample size (with α=.05, power=.95, η^2=0.05), and the final estimate was 331 participants.

A total of 501 participants completed the first part of the survey (ie, performance-based health literacy and sociodemographics). Among them, 380 participants completed the second part of the survey and 121 participants dropped out. Of the 380 participants, 18 were excluded because their answers to the postmanipulation check were incorrect, suggesting deficiencies in their perception of the website being assessed. Comparisons between dropouts and participants who completed the surveys did not show any significant difference in terms of gender, age, educational level, or performance-based health literacy. The final number of participants was 362. The subjects were aged between 18 and 66 years (mean age 35.54 years, SD 13.76 years) and 72.1% (261/362) were female. The majority of the participants had a high school degree (145/362, 40.1%), a bachelor's degree (73/362, 20.2%), or a master's degree (134/362, 37.0%).

Statistical Analysis

SPSS Statistics for Windows (version 25.0; IBM Corp) and Hayes' PROCESS macro v3.4 for SPSS 25.0 [47] were applied. The internal reliability and consistency of the continuous scales were calculated using α and r_s, and the Kuder-Richardson 20 coefficient was used for noncontinuous scales. Normality distribution was found with skewness values ranging between –1.1 and 0.44 and kurtosis values ranging between 0.9 and –0.62, except for the variable assessing the information-seeking behavior about depression (skewness 2.5, kurtosis 6.6). Chi-square tests and *t* tests were calculated to compare the subsample of participants who completed only the first part of the survey with those who completed the entire survey. Pearson correlations were calculated between continuous variables and point-biserial correlation coefficients (r_{pb}) were calculated between binary and continuous variables. We conducted *t* tests to compare the mean differences in perceived information quality and the level of health literacy. Logistic regression was performed with treatment choice as the outcome variable and gender, age, past experience with depression, both measures of health literacy, and an interaction term between health literacy and perception of website quality as independent variables. Using the entry method, we ran a three-step model, including the relevant sociodemographic variables first, then the possible predictors of the main effects on treatment choice, and lastly the interaction term. Finally, mediation analyses were carried out on perceived information quality as a mediator between health literacy and treatment choice.

Results

Bivariate Findings

The two measures of health literacy were not associated (r=0.06, P=.24, n=362). In response to RQ1, it can be said that the correlation shows that there is little reason to hold that the two scales measure the same thing.

Correlation analysis also showed that respondents with higher (as compared with lower) levels of performance-based health literacy perceived a low-quality website as more negative and were inclined to choose beneficial behaviors for themselves, family, and friends. In contrast, when respondents were grouped according to perception-based health literacy, no such difference appeared. However, those with high health literacy levels rated a high-quality website as more positive than those with low levels did. This supports H1 and also H2, but only for the judgment of low-quality websites. It also supports H4, except for positive websites, which were perceived more positively by individuals who perceived themselves as highly health literate and for whom no such association was expected. Correlations between health literacies and dependent variables are displayed in Table 1.

Table 1. Correlations between two health literacy measures and perception of website quality/choice of depression treatment.a

	Objective, performance-based measure (NVS)[b], n=362		Subjective, perception-based measure (eHEALS)[c], n=362	
	r	*P* value	*r*	*P* value
Perception of high-quality website	0.02	.78	0.14	.009
Perception of low-quality website	–0.14	.006	–0.08	.13
Treatment preferences	0.12	.007	0.01	.81

[a] High numbers indicate high health literacy, positive perception, and beneficial preferences for treatment.

[b] NVS: Newest Vital Sign.

[c] eHEALS: eHealth Literacy Scale.

An independent samples *t* test was conducted to evaluate whether the performance-based health literacy level (low versus high) had an effect on the perceived information quality of the low-quality website. The analysis was significant ($t(360)=2.81$, $P=.005$), with participants who had higher performance-based health literacy achieving lower scores (mean 3.70, SD 1.53) than participants who had lower performance-based health literacy (mean 4.17, SD 1.61). The same analysis was conducted considering self-reported health literacy, but the result was not significant ($t(360)=1.064$, $P=.29$).

In the next step, we tried to build linear regression models for predicting the perception of high- and low-quality mental health websites and treatment choice. We expected respondents with high levels of objective health literacy to evaluate the two websites more consistently than respondents with low levels of health literacy and make a better treatment choice. Consistent evaluation means that the high-quality and low-quality websites were assessed positively and negatively, respectively. A similar difference was not expected to emerge for respondents with different levels of subjective health literacy. Predictors were entered stepwise: the sociodemographic variables (gender, age, education) were entered first, then the covariates indicating experience, and finally the measures of health literacy. Table 2 shows the coefficients of the regression analyses.

Table 2. Regression analyses for predicting the perception of high- and low-quality mental health websites and treatment choice.

Step	Evaluation of high-quality website (linear regression)		Evaluation of low-quality website (linear regression)		Treatment choice (binary logistic regression)	
	β	*P* value	β	*P* value	Exp(B)	*P* value
Step 1: Sociodemographics						
Gender (0=males)	–.056	.30	.035	.497	1.866	.02
Age	.002	.98	.122	.02	0.990	.29
Education level (0=low)	–.036	.50	–.179	<.001	0.998	.99
R^2 (%)	0.3		4.1		4.8	
Step 2: Experience						
Health information–seeking	.004	.95	.028	.63	1.042	.77
Depression-related information–seeking	.023	.69	–.052	.36	1.108	.31
Past experience of depression	–.018	.75	–.128	.02	1.306	.30
R^2 (%)	0.7		5.1		6.2	
Step 3: Health literacy						
Objective, performance-based health literacy (NVS[a])	–.019	.73	–.099	.07	2.068	.004
Subjective, perception-based health literacy (eHEALS[b])	.144	.01	–.041	.47	0.697	.04
R^2 (%)	2.4		5.6		10.4	
Step 4: Perception of mental health websites						
High-quality site	N/A[c]	N/A	N/A	N/A	1.253	.03
Low-quality site	N/A	N/A	N/A	N/A	1.088	.74
R^2 (%)	N/A		N/A		18.6	
Step 5: Interaction						
Interaction between objective health literacy and perception of a low-quality site	N/A		N/A		0.886	.035
R^2 (%)	N/A		N/A		20.1	

[a] NVS: Newest Vital Sign.

[b] eHEALS: eHealth Literacy Scale.

[c] N/A: not applicable.

The first noteworthy result was that the only predictor of a favorable perception of a high-quality website was self-perceived health literacy. A low-quality website was evaluated negatively by highly educated persons, by persons with past experience of depression, and almost (missing significance by a small margin) by objective health literacy. Treatment choice was influenced by gender, as well as by objective health literacy (beneficially) and subjective health literacy (detrimentally). Recognition of a high-quality website by positive perception also increased participants' likelihood of making a sound treatment choice.

The interaction analysis showed that individuals with a high level of objective health literacy and an accurate recognition of a low-quality website as being problematic exercised good judgment regarding treatment behaviors. In particular, participants with high performance-based health literacy were more likely to choose the right treatment for depression than participants with low performance-based health literacy.

The mediation analysis yielded similar results. First, an overall path led from health literacy to treatment choice, but only when health literacy was operationalized as NVS—that is, an objective, performance-based measure. The subjective measure showed no relationship between health literacy and treatment choice. Second, when health literacy was operationalized objectively and respondents were given a low-quality website to assess, they recognized the low-quality website for what it is, and that recognition improved their judgment of treatment alternatives. Third, individuals with high levels of subjective health literacy did not have better insight into the quality of websites and did not assess website qualities differently than individuals with lower health literacy; however, if they recognize the poor quality of the website, they are also capable of opting for better treatments. In other words, individuals with high levels of subjective health literacy were prone to overlook the poor quality of web content, but if they do not overlook it, they will make better treatment decisions (Table 3 and Figure 1).

Table 3. Results of mediation analysis.

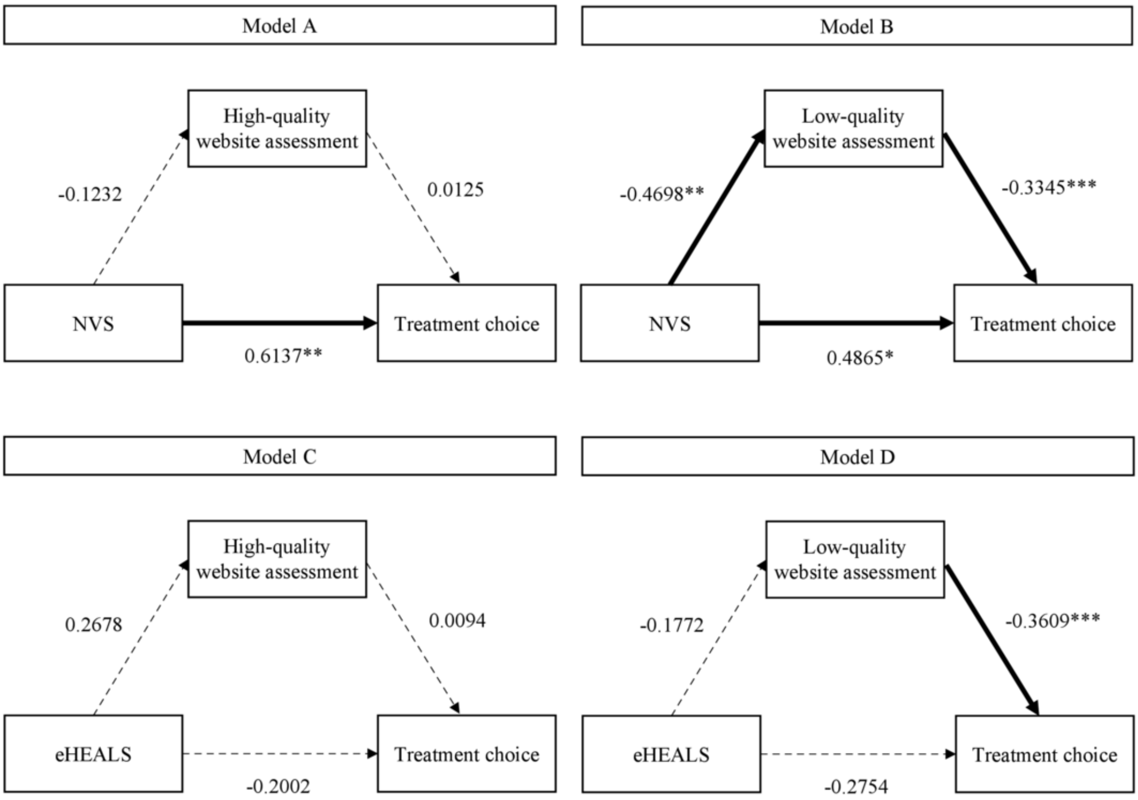
Mediation analysis	Model A	Model B	Model C	Model D
Model specification				
Health literacy indicator	NVS[a]	NVS	eHEALS[b]	eHEALS
Quality of website to be assessed	High	Low	High	Low
Model				
Path from health literacy to website assessment	–0.1232 (*P*=.38)	–0.4698 (*P*=.005)[c]	0.2678 (*P*=.05)	–0.1772 (*P*=.28)
Path from health literacy to treatment choice	0.6137 (*P*=.007)[c]	0.4865 (*P*=.04)[c]	–0.2002 (*P*=.37)	–0.2754 (*P*=.23)
Path from website assessment to treatment choice	0.0125 (*P*=.88)	–0.3345 (*P*<.001)[c]	0.0094 (*P*=.91)	–0.3609 (*P*<.001)[c]
Effect				
Direct effect of health literacy on treatment choice	0.6137 (*P*=.007)[c]	0.4865 (*P*=.04)[c]	–0.2002 (*P*=.37)	0.2754 (*P*=.23)
Indirect effect of health literacy on treatment choice	–0.0015 (*P*=.92)	0.1576 (*P*=.02)[c]	0.0025 (*P*=.92)	0.0640 (*P*=.31)
Model summary				
–2LL	455.2257	434.8030	462.8626	437.7285
P value	.02	<.001	.67	<.001

[a] NVS: Newest Vital Sign.

[b] eHEALS: eHealth Literacy Scale.

[c] Coefficients with *P*<.05.

Figure 1. Mediation analyses. eHEALS: eHealth Literacy Scale; NVS: Newest Vital Sign. **P*<.05; ***P*<.01; ****P*<.001.



Discussion

Association of Performance-Based and Perception-Based Hypotheses (RQ1)

For a summary of the results, we feel that the research question highlights the necessity to differentiate subjective and objective health literacy more clearly, conceptionally as well as in terms of measurement. One finding to support this view is the weak association that was found between the two kinds of information. Another reason to differentiate them is because of the different influences that the two kinds of health literacy had on the study outcomes—the assessment of the quality of internet information (as an intermediate variable) and the choice of treatment behaviors, in line with official recommendations.

Effect on Outcomes (H1 and H4)

In general terms, the answer to RQ1 was that performance-based and perception-based health literacies were uncorrelated, and they played very different parts in predicting the dependent variables. These results and other published research [10] suggest that we need to think seriously about referring to apparently very different concepts by the same name. The study clearly shows that health literacy positively affects patients' choice of treatment, "positively" indicating that high health literacy and beneficial decision making are linked. However, this was true only for objective, performance-based health literacy and not for subjective, perception-based concepts of health literacy. This clearly supports H1 and H4, and it means that the perception-based version of health literacy is lacking the major asset of the concept—improved judgment regarding medical treatments.

Role of Intermediate Variables (H2 to H4)

Health literacy helps patients to identify low-quality health information on the web (H2), and those who recognize websites of low quality tend to make better treatment choices (H3). Both assertions are true for patients whose health literacy is defined objectively; when it is defined subjectively, the link between literacy and quality perception disappears (H4), although the effect of recognizing the problematic site remains. In other words, patients who recognize a low-quality website have better judgment, but a subjective classification of health literacy does not play a role in that recognition.

The objective, factual ability to find one's way through the communicative maze of present-day health care has several components, including the ability to describe one's symptoms to the doctor and understand the doctor's reasoning of diagnosis and treatment. It is something genuinely different from patients' personal impressions that a consultation went well, that they succeeded in presenting their case, and that they understood what the doctor had to say [7].

Some but not all expectations about performance-based health literacy were met in this study. Performance-based literacy predicts the recognition of the shortcomings of low-quality mental-health websites (H2) and good judgment with regard to treatment (H1), and health literacy and accurate perceptions together predict good treatment choices even better (H3). Favorable evaluations of websites, in contrast, seem unrelated to this kind of health literacy, other than what was hypothesized in H2, which pertained to both low- and high-quality websites. Although perception-based health literacy was associated with favorable perceptions of high-quality websites, which went against H4, H4 was otherwise supported.

In the wake of this analysis, health literacy no longer appears to provide the certain benefits that we initially described. However, objective health literacy can still be perceived to be a valuable asset. Its association with sound judgments on treatment and its relationship with negative evaluations of low-quality websites reflect its value. After reviewing our results, the same value cannot be placed on subjective, perception-based health literacy, which might fail to lead to a successful result even if it does not actually do harm.

Actions aimed at enhancing health literacy should not stop at influencing subjective health literacy. The more important point of attack seems to be objective health literacy, and leaving it out of the picture runs the risk of failing to reach the desired outcome. The autonomy policies in health care today are an example of this. They might easily induce in patients a desire to make treatment choices by themselves, and with this growing desire, patients may increasingly consider themselves to be the agency in charge; this might blind individual patients to the fact that nothing much has changed with regard to the objective abilities they possess.

Health literacy on one side and autonomy ideals on the other side complement one another. The abilities that come with health literacy can be used to prevent patients from making too many detrimental decisions, and the autonomy granted would give their health literacy a raison d'être as it would provide a chance to apply their newly acquired ability in the real world, under real circumstances and with real consequences. It has to be noted, though, that health literacy is understood as being of assistance to physicians, not as their replacement.

The desire to counter internet disinformation by increasing health literacy is based on patients' ability to recognize the low quality of some of the messages on the web. According to our results, this can work for objective, performance-based health literacy, but not for the subjective version of the concept. This might be explained by a specific shortcoming of thinking: one is relatively satisfied with their communication abilities because one cannot imagine how much better they could be. This would apply to the ease with which one communicates in health contexts as well as to the quality of websites. Low standards, applied to oneself and to websites as well, could be the common problem behind the measured association.

We expected to find a differential effect on an individual's decision making based on what they know and what they think they know on the given topic [5,10]. In this vein, we expected that self-reported health literacy (ie, what individuals thought they knew) would not make any difference in the depression-related decision making compared with performance-based health literacy (ie, what individuals actually knew). The hypothesis was confirmed because we found that high objective health literacy predicted sound treatment choice, whereas high subjective health literacy did not. Measuring concepts as subjective entities, as is done for intelligence and knowledge [20], goes back to the idea that people react primarily not to objective states but to their perception of those states. Asking people what they think they know, or how intelligent they think they are, works well because these are everyday terms. Health literacy, however, is not an everyday concept, and therefore its measurement must be performed indirectly.

Moreover, subjective and objective health literacy did not correlate. This result was in line with the literature supporting the inadequacy of self-reported measures to detect one's actual health literacy level [48] and demonstrated the fact that the two measures evaluate different constructs [21,39,48,49].

We also tested whether information quality mediated the relation between health literacy and depression-related decision making, first considering performance-based and then self-reported health literacy. The analyses revealed that the perceived quality of a low-quality message partially mediated the relation between the performance-based health literacy level and the treatment choice. Therefore, the performance-based health literacy level not only had a positive direct effect on treatment choice but it also had a positive effect on treatment choice through its effect on the perceived quality in the case of a low-quality message. In other words, the higher the actual health literacy level, the lower the perceived quality of a low-quality website, and the better the treatment choice, while the lower the actual health literacy, the higher the perceived quality of a low-quality message, and the worse the decision making. This finding is meaningful because it suggests that efforts to increase the objective health literacy level have the potential to improve not only the judgment of information quality but also the appropriateness of the decision making.

Moreover, in the various mediation models examined, the main effect of performance-based health literacy on treatment choice was always significant (H1), while the main effect of self-reported health literacy was not (in any of the cases [H4]). This result suggests that the self-reported health literacy level does not have the potential to improve the quality of judgments and decisions.

To sum up, believing oneself to be health literate does not imply that one is able to make accurate perceptions of the quality of health information presented to them, nor does it mean that one is able to make good health-related decisions. Feeling empowered, indeed, was argued not to be a sufficient requisite for making good decisions [5,10]. These findings provide further support to the hypothesis that a difference actually exists between performance-based and self-reported measures of health literacy. Interpreting self-confidence as health literacy does not suffice; the more consequential performance-based level of health literacy counts more [21,29,48,49], especially in terms of making accurate judgments of health information.

Strengths and Limitations

One main limitation of the present study is the generalizability of its results. We recognize that there might have been a selection bias due to the sampling method used; however, a positive aspect was that the sampling was conducted via Facebook. Therefore, we were very likely to sample actual internet users. Another limit to the generalizability of our findings pertains to the ecological validity of the study. Indeed, we selected the mock versions of two actual websites as experimental stimuli. However, in the future, better selection criteria could be established by reviewing the main Italian information websites about depression using appropriate tools, such as DISCERN [50].

Strictly speaking, the analysis cannot claim to say anything beyond the findings on mental health and depression websites and their users. The results were nevertheless formulated without constant reminders of this limitation. This can be interpreted as a hint that we expect to find similar associations for other health topics as well.

Conclusions

The take-home messages of the present research are as follows. Our findings provide further support to the hypothesis that a difference exists between performance-based and self-reported measures of health literacy. In particular, performance-based health literacy is associated with the ability to recognize faulty information, sound judgment and health-related decision-making skills, and good health outcomes. These are largely the consequences that make health literacy such a cherished concept. On the other hand, individuals who boast of experience that might indicate health literacy (ie, the items in the subjective, self-reported measures) cannot be regarded as having similar abilities, especially in the absence of objective health literacy to support their conception of their own abilities. Simply because someone believes that they communicate well in health care environments does not mean they actually do and no one should assume that they do. The state of not knowing what one does not know has several consequences. First, other individuals in the health care system have to assume that, despite the professed communicative abilities, these individuals are prone to become victims of medical disinformation. Second, some questions arise. For performative measures, research should ask how much health literacy is needed to avoid adverse decisions in health care. For self-reported literacy measures, the poor construct validity for health literacy suggests that these measures need to be tested directly to determine whether they are valid measures of empowerment [10].

In terms of health policies, the conclusion is that patients should not be encouraged to claim a larger share in medical decision making unless it is clear that a sufficient basis of objective health literacy exists.

Acknowledgments

PJS conceived the project and wrote, together with UH, the final version of the paper. AP collected data as a part of her Master's thesis. PJS and SP supervised AP's work and performed the analyses, and SP wrote the first draft of the paper. All authors read and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pilot studies.
[DOCX File , 18 KB - jmir_v23i1e20457_app1.docx]

Multimedia Appendix 2
Websites.
[DOCX File , 245 KB - jmir_v23i1e20457_app2.docx]

References

1. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Prom Int 2000;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
2. Engel G. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 08;196(4286):129-136. [doi: 10.1126/science.847460] [Medline: 847460]
3. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. JAMA 1992;267(16):2221-2226. [Medline: 1556799]
4. Ong L, de Haes J, Hoos A, Lammes F. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [doi: 10.1016/0277-9536(94)00155-m] [Medline: 7792630]
5. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
6. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
7. Ishikawa H, Yano E. Patient health literacy and participation in the health-care process. Health Expect 2008 Jun;11(2):113-122 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00497.x] [Medline: 18494956]
8. Paasche-Orlow MK, Wilson EAH, McCormack L. The evolving field of health literacy research. J Health Commun 2010;15 Suppl 2:5-8. [doi: 10.1080/10810730.2010.499995] [Medline: 20845188]
9. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
10. Schulz PJ, Nakamoto K. Health literacy and patient empowerment in health communication: the importance of separating conjoined twins. Patient Educ Couns 2013 Jan;90(1):4-11. [doi: 10.1016/j.pec.2012.09.006] [Medline: 23063359]
11. Kiechle ES, Bailey SC, Hedlund LA, Viera AJ, Sheridan SL. Different Measures, Different Outcomes? A Systematic Review of Performance-Based versus Self-Reported Measures of Health Literacy and Numeracy. J Gen Intern Med 2015 Oct;30(10):1538-1546 [FREE Full text] [doi: 10.1007/s11606-015-3288-4] [Medline: 25917656]
12. Nguyen TH, Paasche-Orlow MK, McCormack LA. The State of the Science of Health Literacy Measurement. Stud Health Technol Inform 2017;240:17-33 [FREE Full text] [Medline: 28972507]
13. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
14. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
15. Burkhanov SA, Dormeneva EV, Kosykh VA, Berdichevskiĭ VR, Torchilin VP. [Interaction of liposomes with a varying lipid composition with hepatocytes in vitro]. Biull Eksp Biol Med 1985 Jun;99(6):679-681. [Medline: 4016258]
16. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, HLS-EU Consortium. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 2013 Oct 10;13:948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
17. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995 Oct;10(10):537-541. [doi: 10.1007/BF02640361] [Medline: 8576769]
18. Woloshin S, Schwartz LM, Welch HG. Patients and medical statistics. Interest, confidence, and ability. J Gen Intern Med 2005 Nov;20(11):996-1000 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00179.x] [Medline: 16307623]
19. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making 2001;21(1):37-44. [doi: 10.1177/0272989X0102100105] [Medline: 11206945]
20. Cianciolo AT, Grigorenko EL, Jarvin L, Gil G, Drebot ME, Sternberg RJ. Practical intelligence and tacit knowledge: Advancements in the measurement of developing expertise. Learning and Individual Differences 2006 Jan;16(3):235-253. [doi: 10.1016/j.lindif.2006.04.002]

21. Mantwill S, Allam A, Camerini A, Schulz PJ. Validity of Three Brief Health Literacy Screeners to Measure Functional Health Literacy - Evidence from Five Different Countries. J Health Commun 2018;23(2):153-161. [doi: 10.1080/10810730.2017.1417515] [Medline: 29319424]
22. Smith SG, Curtis LM, O'Conor R, Federman AD, Wolf MS. ABCs or 123s? The independent contributions of literacy and numeracy skills on health task performance among older adults. Patient Educ Couns 2015 Aug;98(8):991-997 [FREE Full text] [doi: 10.1016/j.pec.2015.04.007] [Medline: 25936579]
23. Gazmararian JA, Kripalani S, Miller MJ, Echt KV, Ren J, Rask K. Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. J Gen Intern Med 2006 Dec;21(12):1215-1221 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00591.x] [Medline: 17105519]
24. Kobayashi LC, Smith SG. Cancer Fatalism, Literacy, and Cancer Information Seeking in the American Public. Health Educ Behav 2016 Aug;43(4):461-470 [FREE Full text] [doi: 10.1177/1090198115604616] [Medline: 26377524]
25. van der Heide I, Uiters E, Rademakers J, Struijs JN, Schuit AJ, Baan CA. Associations among health literacy, diabetes knowledge, and self-management behavior in adults with diabetes: results of a dutch cross-sectional study. J Health Commun 2014;19 Suppl 2:115-131. [doi: 10.1080/10810730.2014.936989] [Medline: 25315588]
26. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
27. Wångdahl J, Lytsy P, Mårtensson L, Westerling R. Poor health and refraining from seeking healthcare are associated with comprehensive health literacy among refugees: a Swedish cross-sectional study. Int J Public Health 2018 Apr;63(3):409-419 [FREE Full text] [doi: 10.1007/s00038-017-1074-2] [Medline: 29480325]
28. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun 2012;17 Suppl 3:141-159. [doi: 10.1080/10810730.2012.712615] [Medline: 23030567]
29. Hirsh JM, Boyle DJ, Collier DH, Oxenfeld AJ, Nash A, Quinzanos I, et al. Limited health literacy is a common finding in a public health hospital's rheumatology clinic and is predictive of disease severity. J Clin Rheumatol 2011 Aug;17(5):236-241. [doi: 10.1097/RHU.0b013e318226a01f] [Medline: 21778910]
30. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
31. Borzekowski DLG, Rickert VI. Adolescent cybersurfing for health information: a new resource that crosses barriers. Arch Pediatr Adolesc Med 2001 Jul;155(7):813-817. [doi: 10.1001/archpedi.155.7.813] [Medline: 11434849]
32. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. J Med Internet Res 2004 Mar 01;6(1):e7 [FREE Full text] [doi: 10.2196/jmir.6.1.e7] [Medline: 15111273]
33. Helft PR, Eckles RE, Johnson-Calley CS, Daugherty CK. Use of the internet to obtain cancer information among cancer patients at an urban county hospital. J Clin Oncol 2005 Aug 01;23(22):4954-4962. [doi: 10.1200/JCO.2005.09.621] [Medline: 16051948]
34. Gauld R, Williams S. Use of the Internet for health information: a study of Australians and New Zealanders. Inform Health Soc Care 2009 Sep;34(3):149-158. [doi: 10.1080/17538150903102448] [Medline: 19670005]
35. Richter JG, Becker A, Koch T, Nixdorf M, Schacher B, Monser R, et al. Internet use in rheumatology outpatients in 2006: gender less important. Clin Exp Rheumatol 2009;27(1):15-21. [Medline: 19327224]
36. Feufel MA, Stahl SF. What do web-use skill differences imply for online health information searches? J Med Internet Res 2012 Jun 13;14(3):e87 [FREE Full text] [doi: 10.2196/jmir.2051] [Medline: 22695686]
37. Benotsch EG, Kalichman S, Weinhardt LS. HIV-AIDS patients' evaluation of health information on the internet: the digital divide and vulnerability to fraudulent claims. J Consult Clin Psychol 2004 Dec;72(6):1004-1011. [doi: 10.1037/0022-006X.72.6.1004] [Medline: 15612847]
38. Capecchi L, Guazzini A, Lorini C, Santomauro F, Bonaccorsi G. The first italian validation of the most widespread health literacy assessment tool: the Newest Vital Sign. Epidemiol Prev 2015;39(4 Suppl 1):124-128 [FREE Full text] [Medline: 26499429]
39. Diviani N, Dima AL, Schulz PJ. A Psychometric Analysis of the Italian Version of the eHealth Literacy Scale Using Item Response and Classical Test Theory Methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
40. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
41. European Alliance Against Depression. https://ifightdepression.com/it/. Essere Informati. URL: https://ifightdepression.com/it/ [accessed 2020-05-04]
42. La Mente è Meravigliosa. I trattamenti più efficaci contro la depressione. La Mente è Meravigliosa. URL: https://lamenteemeravigliosa.it/i-trattamenti-piu-efficaci-contro-la-depressione/ [accessed 2020-05-04]
43. Sanz J, Garcia-Vera MP. Misconceptions about depression and its treatment (II). Pap. Psicol 2017;38(3):169-176. [doi: 10.23923/pap.psicol2017.2833]

44. Freeman JL, Caldwell PHY, Bennett PA, Scott KM. How Adolescents Search for and Appraise Online Health Information: A Systematic Review. J Pediatr 2018 Apr;195:244-255.e1. [doi: 10.1016/j.jpeds.2017.11.031] [Medline: 29398062]
45. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 2015 Dec 11;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
46. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
47. Hayes A. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. New York, NY: The Guilford Press; 2013.
48. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
49. Gerich J, Moosbrugger R. Subjective Estimation of Health Literacy-What Is Measured by the HLS-EU Scale and How Is It Linked to Empowerment? Health Commun 2018 Mar;33(3):254-263. [doi: 10.1080/10410236.2016.1255846] [Medline: 28033479]
50. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]

Abbreviations

BHLS: Brief Health Literacy Screen
eHeals: eLiteracy scale
HLS-EU-Q47: European Health Literacy Survey Questionnaire
NVS: Newest Vital Sign
REALM: Rapid Estimate of Adult Literacy in Medicine
TOFHLA: Test of Functional Health Literacy in Adults

Edited by G Eysenbach, R Kukafka; submitted 20.05.20; peer-reviewed by K Read, S McLennan; comments to author 15.06.20; revised version received 24.07.20; accepted 03.11.20; published 21.01.21.

Please cite as:
Schulz PJ, Pessina A, Hartung U, Petrocchi S
Effects of Objective and Subjective Health Literacy on Patients' Accurate Judgment of Health Information and Decision-Making Ability: Survey Study
J Med Internet Res 2021;23(1):e20457
URL: http://www.jmir.org/2021/1/e20457/
doi:10.2196/20457
PMID:33475519

Original Paper

Role of Health Literacy in Health-Related Information-Seeking Behavior Online: Cross-sectional Study

Hee Yun Lee[1], MSW, PhD; Seok Won Jin[2], PhD; Carrie Henning-Smith[3], PhD; Jongwook Lee[4], PhD; Jaegoo Lee[5], PhD

[1]School of Social Work, The University of Alabama, Tuscaloosa, AL, United States
[2]School of Social Work, University of Memphis, Memphis, TN, United States
[3]School of Public Health, University of Minnesota, Minneapolis, MN, United States
[4]Department of Global Health and Population, Harvard T H Chan School of Public Health, Harvard University, Boston, MA, United States
[5]School of Social Work, Jackson State University, Jackson, MS, United States

Corresponding Author:
Jongwook Lee, PhD
Department of Global Health and Population
Harvard T H Chan School of Public Health
Harvard University
677 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 617 432 1232
Email: jongwook_lee@hsph.harvard.edu

Abstract

Background: The internet has emerged as a main venue of health information delivery and health-related activities. However, few studies have examined how health literacy determines online health-related behavior.

Objective: The aim of this study was to investigate the current level of health-related information-seeking using the internet and how health literacy, access to technology, and sociodemographic characteristics impact health-related information-seeking behavior.

Methods: We conducted a cross-sectional study through a survey with Minnesotan adults (N=614) to examine their health literacy, access to technology, and health-related information-seeking internet use. We used multivariate regression analysis to assess the relationship between health-related information-seeking on the internet and health literacy and access to technology, controlling for sociodemographic characteristics.

Results: Better health literacy (β=.35, SE 0.12) and greater access to technological devices (eg, mobile phone and computer or tablet PC; β=.06, SE 0.19) were both associated with more health-related information-seeking behavior on the internet after adjusting for all other sociodemographic characteristics. Possession of a graduate degree (β=.28, SE 0.07), female gender (β=.15, SE 0.05), poor health (β=.22, SE 0.06), participation in social groups (β=.13, SE 0.05), and having an annual health exam (β=.35, SE 0.12) were all associated with online health-related information-seeking.

Conclusions: Our findings indicate that access to online health-related information is not uniformly distributed throughout the population, which may exacerbate disparities in health and health care. Research, policy, and practice attention are needed to address the disparities in access to health information as well as to ensure the quality of the information and improve health literacy.

(J Med Internet Res 2021;23(1):e14088) doi:10.2196/14088

KEYWORDS

digital divide; health literacy; internet; technology; access

Introduction

Health information access and use are essential for optimal health outcomes [1]. For example, a meta-analysis found that health information access and use were associated with better compliance with medical treatment in patients with chronic and acute illnesses [2]. Another study demonstrated that health information access and use are associated with self-care behavior among patients with heart failure [3]. Furthermore, several previous studies revealed that health information access and use predict other health-related behaviors, including having an annual medical checkup and undergoing cancer screenings, because increased cognitive knowledge in health can lead to behavioral outcomes such as receiving preventive screenings [4-6]. This association shows that securing adequate access to and use of health information plays a key role in improving health outcomes in varied domains.

Since its advent, the internet has served as a primary medium to convey health information [7,8]. For the last few decades, the dramatic increase in the usage of internet-embedded devices such as laptops, tablets, and smartphones has enabled people to access and use health information at any place and time [9]. According to recent national surveys [10,11], in the United States, approximately three-quarters of adults have broadband internet service at home, nearly 90% use the internet, nearly two-thirds of adults own a smartphone, and nearly 80% use the internet for various health-related purposes such as seeking health information, communicating with doctors, and purchasing medicine or vitamins [12-14].

Although the internet has contributed to enhancing the potential for health information access and use while reducing barriers to obtaining health information in the US general population, not all people have obtained these benefits. For example, gender is a key predictor of use of the internet and online health information behavior [15]. Previous studies have shown that men are less likely than women to use the internet for health information–seeking and to trust online sources [13,16-18]. There remain some groups of people who still experience difficulties in accessing and utilizing health information because of barriers such as a digital gap and limited health literacy [1,19,20].

A digital gap is defined as unequal access to technology and capability of its usage [21]. These variations in access to and use of technology often lead to disparities in health outcomes [22]. Existing research shows that predictors of the digital gap include age, gender, education attainment, and income [20]. Closing the digital gap is important because it is significantly associated with the enhancement of health-related decision making, health behavior, and health care system navigation [23,24]. In addition to the digital gap, health literacy also influences health information access and use. Health literacy is defined as “the degree to which individuals can obtain, process, understand, and communicate about health-related information needed to make informed health decisions” [25]. Researchers agree on the importance of health literacy in that people with low health literacy are at a high risk of varied poor health outcomes [26-29].

Despite consistent emphasis on the importance of health literacy as a determinant of health outcomes [30,31], and the emergence of the internet as a main venue of health information delivery and health-related activities, few studies have examined how health literacy determines online health-related behaviors such as the usage of the internet for health information–seeking and health-related activities, including online scheduling for visits to clinics. Therefore, this study had a threefold purpose: (1) to investigate the levels of online health-related behavior, (2) to assess whether health literacy is associated with online health-related behavior, and (3) to examine whether sociodemographic characteristics are associated with online health-related behavior. The findings of this study will offer insights into health literacy and sociodemographic-specific interventions for improving online health-related behaviors and health outcomes.

Methods

Research Design and Data Collection

The research team collected survey data from 732 adults aged 18 years or older at the 2016 Minnesota State Fair with approval from the University of Minnesota Institutional Review Board. The survey included questions about health literacy and health behavior in addition to general sociodemographic information. Participants received a small gift (a backpack with the University of Minnesota logo, which is worth about US $3) after they completed their voluntary survey using REDCap software via an iPad. Owing to missing values, data from 614 respondents (241 men and 373 women) out of the original 732 were used in the study as the analytic sample. Comparing the included analytic sample of 614/732 (83.9%) to the excluded sample of 118/732 (16.1%) due to missing values, we found that people in the included sample, who completed the survey, were more likely to have a family cancer history (included sample: 480/614, 78.2%; excluded sample: 49/76, 64.5%; P=.02), have both a smartphone and a computer or tablet PC (included sample: 485/614, 79.0%; excluded sample: 36/63, 57.1%; P<.001), and use internet for health information (included sample: 549/614, 89.4%; excluded sample: 57/74, 77.0%; P=.02) at the 5% significance level.

Instruments and Measures

To measure health-related information-seeking behavior, the main outcome variables of this study were based on two sets of questions about seeking health-related information on the internet. First, a survey question asked whether the respondent used the internet for health information, which was used as a binary outcome variable for health-related internet use. Second, the total score of the 12 health-related internet use items from the Health Information National Trends Survey [32] questionnaire was used to measure the participants’ health-related information-seeking internet use score. Each question was assigned a value of 0 (no) or 1 (yes). The Cronbach α of the 12 items for the health-related internet use score was .71. Higher health-related internet use scores indicate more health-related information-seeking behavior using the internet.

The two key independent variables were health literacy and access to technology devices. Health literacy is comprised of the total score of the following three health literacy items from the Behavioral Risk Factor Surveillance System Questionnaire [33] developed by the US Centers for Disease Control and Prevention, and was used as the primary measure of health literacy in this study: (1) “How difficult is it for you to get advice or information about health or medical topics if you needed it?” (2) “How difficult is it for you to understand information that doctors, nurses, and other health professionals tell you?” and (3) “You can find written information about health on the internet, in newspapers and magazines, and in brochures in the doctor’s office and clinic; in general, how difficult is it for you to understand written health information?” Each health literacy item was measured by a 5-point Likert scale ranging from 1 (“not at all”) to 5 (“always”), and the Cronbach α of the three items was .74. Access to technology devices was measured by a categorical variable assessing the type(s) of devices the respondent had access to. Respondents chose one of the following options: “no device,” “computer or tablet PC at home,” “mobile phone,” “mobile phone and computer or tablet PC at home,” “smartphone,” or “smartphone and computer or tablet PC at home.”

In addition to the measures on health literacy and technology devices, the sociodemographic and health information of participants, including gender, age, marital status (never married vs married/partnered), educational level (less than a bachelor’s degree, bachelor’s degree, or graduate degree and higher), annual health checkup in the past 12 months, any family cancer history, health status (poor/fair vs good/very good/excellent), and participation in a social group, were included as covariates.

Data Analysis

We first investigated the association of health-related information-seeking internet use with sociodemographic characteristics using *t* tests for binary variables and *F* tests for categorical variables with more than two values. We report the Pearson correlation coefficients (*r*) for continuous variables. In particular, we focused on gender differences in health-related information-seeking internet use because previous studies have shown gender differences in internet use and health information–seeking behavior [13,15]. Next, we used multiple regression analyses for both binary and continuous outcome values of health-related internet use. We used logistic regression analysis for binary outcomes and ordinary least-squares regression analysis for the continuous health-related internet use score. We used heteroscedasticity robust standard errors for the multiple regression analyses. We conducted all analyses in Stata 14.1, using a 5% statistical significance level criterion.

Results

Sociodemographic Characteristics and Bivariate Analysis

Table 1 summarizes the sociodemographic characteristics of the study sample. A total sample of 614 was used in the study, with a majority of women. The mean health-related information-seeking internet use score was significantly higher for women than for men. The mean age of the sample was 41.87 years (SD 16.83), and the correlation between age and health-related internet use score was negligible (r=0.025, P=.54). Only 408 of the 614 participants (66.4%) in the sample reported their race/ethnicity.

The total health-related internet use score was significantly different among educational level groups, which was the lowest among individuals with less than a bachelor’s degree and was the highest among individuals with a graduate degree. This demonstrates that participants with higher education had higher health-related information-seeking internet use. The majority of respondents indicated having gone for an annual health checkup, and their health-related internet use score was significantly higher than that for those who had not had an annual health checkup. The majority of the sample reported that their health status was good, very good, or excellent; participants with a lower self-reported health status had higher health-related internet use scores than those with a higher self-reported health status.

The majority of the sample had a smartphone and a computer or tablet PC at home, followed by those with a smartphone only. The health-related internet use scores across the different technology device possession groups were significantly different, with the highest scores for those using a smartphone and a computer or tablet PC, followed by the scores for those who possessed a smartphone only. The mean score for those who owned a mobile phone and computer or tablet PC at home was higher than that of respondents who owned a computer or tablet PC only at home or those who had only a mobile phone. The score for those who did not own any technology device was the lowest, as expected. The mean value of the health literacy total score was 12.37 (SD 2.41, range 0-15), and the Pearson correlation coefficient with the health-related internet use score was 0.079, which is moderate (P=.04).

Table 1. Sociodemographic characteristics and their relation to health-related internet use of the study sample (N=614).

Characteristics[a]	Value	Health-related internet use[b]		
		Mean (SD)	Test statistic[c]	*P* value
Gender, n (%)			$t_{489.585}=2.95$	.003
Male	241 (39.3)	3.27 (2.49)		
Female	373 (60.7)	3.87 (2.35)		
Marital status, n (%)			$t_{610.169}=1.41$	.16
Never married or other	315 (51.3)	3.50 (2.54)		
Married or partnered	299 (48.7)	3.78 (2.28)		
Educational level, n (%)			$F_{2,611}=14.28$	<.001
Less than bachelor’s degree	197 (32.1)	3.06 (2.50)		
Bachelor’s degree	262 (42.7)	3.60 (2.31)		
Graduate degree	155 (25.2)	4.42 (2.30)		
Annual health checkup, n (%)			$t_{277.588}=4.31$	<.001
No	147 (23.9)	2.95 (2.14)		
Yes	467 (76.1)	3.85 (2.46)		
Family cancer history, n (%)			$t_{200.472}=1.70$	.09
No	134 (21.8)	3.31 (2.57)		
Yes	480 (78.2)	3.73 (2.37)		
Self-reported health status, n (%)			$t_{135.037}=2.81$	.006
Very poor/poor/fair	106 (17.3)	4.32 (2.85)		
Good/very good /excellent	508 (82.7)	3.49 (2.30)		
Participating in a social group, n (%)			$t_{611.881}=1.81$	.07
No	321 (52.3)	3.47 (2.53)		
Yes	293 (47.7)	3.82 (2.28)		
Technology devices, n (%)			$F_{5608}=6.28$	<.001
No device	10 (1.6)	0.80 (1.23)		
Computer or tablet PC	7 (1.1)	2.57 (2.07)		
Mobile phone	9 (1.5)	2.00 (1.80)		
Mobile phone + computer or tablet PC	35 (5.7)	2.86 (2.05)		
Smartphone	68 (11.1)	3.13 (2.06)		
Smartphone + computer or tablet PC	485 (79.0)	3.87 (2.46)		

[a] The total sample size of each variable may not be the same as the total sample size of the study due to missing values.

[b] Based on health-related internet use total score (range 0-12).

[c] Two-tailed *t* test assuming unequal variances with Satterthwaite degrees of freedom for binary variables, and *F* test for categorical variables with more than two values.

Table 2 shows the results of the descriptive analysis on internet use for health information and health-related information-seeking internet use items. For the question asking about internet use for health information, 89.4% (549/614) of the sample reported that they have used the internet to look for health or medical information for themselves, with a significant gender difference. Among the 12 health-related internet use items, 6 items were significantly different between men and women at the 5% significance level; women used the internet more for these 6 items, which included “used email or the internet to communicate with a doctor or doctor’s office”; “used a website to help you with your diet, weight, or physical activity”; “looked for a health care provider”; “visited a social networking site such as Facebook or LinkedIn to read and share about medical topics”; “kept track of personal health information such as care received, test results, or upcoming medical appointments”; and “looked for health or medical information for someone else.”

Table 2. Descriptive analysis on health-related internet use.

Question	Total (N=614), mean (SD)	Males (n=241), mean (SD)	Females (n=373), mean (SD)	*t* statistic (df)[a]	*P* value
Internet use for health information (Yes=1, No=0)					
In the past 12 months, have you used the internet to look for health or medical information for yourself? (N=614)	0.89 (0.30)	0.83 (0.37)	0.93 (0.25)	3.63 (379.481)	<.001
Health-related internet use items[b] (Yes=1, No=0)					
Looked for information about quitting smoking (N=610)	0.05 (0.21)	0.06 (0.23)	0.04 (0.19)	0.97 (447.531)	.33
Bought medicine or vitamins online (N=610)	0.18 (0.38)	0.18 (0.39)	0.18 (0.39)	0.11 (504.817)	.91
Participated in an online support group for people with a similar health or medical issue (N=609)	0.06 (0.24)	0.054 (0.23)	0.06 (0.25)	0.65 (540.24)	.52
Used email or the internet to communicate with a doctor or doctor's office (N=611)	0.46 (0.50)	0.39 (0.49)	0.51 (0.50)	2.91 (516.579)	.004
Used a website to help you with your diet, weight, or physical activity (N=611)	0.52 (0.49)	0.47 (0.50)	0.55 (0.49)	2.02 (505.537)	.04
Looked for a health care provider (N=608)	0.42 (0.49)	0.36 (0.48)	0.47 (0.50)	2.75 (519.359)	.006
Downloaded health-related information to a mobile device such as an MP3 player, cell phone, tablet computer, or electronic book device (eg, download mobile apps) (N=610)	0.28 (0.45)	0.30 (0.46)	0.27 (0.45)	0.64 (495.534)	.52
Visited a social networking site such as Facebook or LinkedIn to read and share about medical topics (N=611)	0.29 (0.45)	0.24 (0.43)	0.32 (0.47)	2.23 (537.405)	.03
Wrote in an online diary or "blog" (ie, web log) about any type of health topic (N=605)	0.04 (0.19)	0.03 (0.18)	0.04 (0.20)	0.58 (538.839)	.56
Kept track of personal health information such as care received, test results, or upcoming medical appointments (N=606)	0.43 (0.50)	0.37 (0.48)	0.47 (0.50)	2.59 (518.015)	.01
Looked for health or medical information for someone else (N=599)	0.52 (0.49)	0.44 (0.49)	0.57 (0.49)	3.23 (497.276)	.001
Done anything else health-related on the internet (N=610)	0.39 (0.48)	0.40 (0.49)	0.38 (0.49)	0.60 (506.309)	.55

[a] Two-tailed *t* test assuming unequal variances with Satterthwaite degrees of freedom.

[b] Cronbach α=.708.

Multiple Regression Analyses

Table 3 shows the results of the logistic regression analysis that examined the association of health-related information-seeking internet use with health literacy, sociodemographic factors, and other health-related factors. Women were more likely to use the internet for health-related information-seeking, controlling for other factors, consistent with the bivariate gender comparison for the health-related internet use score (Table 1). Participants who had a postgraduate degree were more likely to use the internet for health-related information-seeking than those who had less than a bachelor's degree. Respondents who had participated in a social group were more likely to use the internet for health information than those who were not in a social group. Respondents with a higher health literacy score were more likely to use the internet for health information. Compared with those who did not have a technology device, those who had a mobile phone and computer or tablet PC at home, those who had a smartphone only, and those who owned a smartphone and computer or tablet PC at home were more likely to use the internet for health-related information. The Wald χ^2 test (Wald χ^2_{17}=53.30, $P<.001$) and pseudo R^2 (0.15) indicated a good model fit to the data.

Table 3. Multiple regression analyses for factors associated with health-related internet use for binary and continuous outcomes (N=614).[a]

Variable	Internet use for health information[b]		Health-related internet use total score[c]	
	Odds ratio (95% CI)	*P* value	Regression coefficient (SE)	*P* value
Female (Reference: Male)	2.68 (1.43-5.00)	.002	0.15 (0.05)	.007
Age	1.40 (0.95-2.07)	.09	0.11 (0.03)	<.001
Age^2	0.99 (0.98-1.00)	.14	–0.002 (0.00)	<.001
Age^3	1.00 (1.00-1.00)	.22	0.00001 (0.00)	.002
Married or partnered (Reference: Not married or partnered)	0.84 (0.41-1.71)	.63	–0.04 (0.05)	.46
Education level				
Less than bachelor’s degree	Reference	N/A[d]	Reference	N/A
Bachelor’s degree	1.19 (0.61-2.34)	.61	0.08 (0.06)	.18
Graduate degree	2.99 (1.15-7.73)	.02	0.28 (0.07)	<.001
Annual health checkup (Reference: No annual health checkup)	1.19 (0.61-2.32)	.61	0.147 (0.06)	.02
Family cancer history (Reference: No family cancer history)	1.19 (0.63-2.27)	.59	0.07 (0.06)	.21
Good/very good/excellent health (Reference: Very bad/bad/fair)	0.61 (0.27-1.40)	.24	–0.22 (0.06)	<.001
Participating in a social group (Reference: not participating)	1.90 (1.03-3.52)	.04	0.13 (0.05)	.008
Log of health literacy total score	7.19 (2.07-25.02)	.002	0.35 (0.12)	.004
Technology devices				
No device	Reference	N/A	Reference	N/A
Computer or tablet PC	4.28 (0.27-68.49)	.31	0.43 (0.31)	.17
Mobile phone	1.89 (0.16-21.81)	.61	0.42 (0.30)	.17
Mobile phone + computer or tablet PC	13.37 (1.53-117.09)	.02	0.60 (0.19)	.002
Smartphone	9.77 (1.48-64.61)	.02	0.73 (0.18)	<.001
Smartphone + computer or tablet PC	8.13 (1.49-44.44)	.02	0.81 (0.17)	<.001

[a] Heterogeneity robust standard errors are used.

[b] Logistic regression for the dichotomous health-related internet use variable.

[c] Ordinary least-squares regression for the continuous health-related internet use total score; the natural logarithm of the health-related internet use score was used as the dependent variable.

[d] N/A: not applicable.

Table 3 also presents the results from the multiple ordinary least-squares regression analysis that investigated the association of the total health-related internet use score with health literacy, health-related factors, and sociodemographic factors. The health-related internet use score and the health literacy score were transformed using the natural logarithm function, which allowed us to interpret the results in approximate percentage changes in the analysis. Women were also more likely (by 14.50%) to have a higher health-related internet use score than men. Age was not significantly associated with the health-related internet use score when we added the age term only; however, age was significantly associated (at a 1% threshold) with the health-related internet use score when we included quadratic and cubic terms of age in addition to the linear term. This means that age was significantly associated with the health-related internet use score in a cubic manner, whereas there was no linear relationship between age and the health-related internet use score. Solving the cubic equation, we found that the health-related internet use score increased with age until about 38 years, decreased between the ages of 38 and 71 years, and increased again after the age of 71 years.

Respondents who had a postgraduate degree had a nearly 27.5% higher health-related internet use score than those who had less than a bachelor’s degree. Participants who had gone for an annual health checkup in the last 12 months had a nearly 14.7% higher health-related internet use score on average compared with those who had not gone for an annual health checkup. People who reported that their general health status was poor or fair had about a 22.6% higher health-related internet use score than those who reported their health status as good, very good, or excellent. The respondents who had participated in a social group had a 12.5% higher health-related internet use score than that of those who do not participate in a social group. People with a 10% higher health literacy score had a 3.5% higher health-related internet use score on average. In addition, health-related internet use scores for people who owned a computer or tablet PC or a mobile phone only were not significantly different from the scores of those without any kind of technology device. By contrast, compared with those without any technology device, people who possessed a mobile phone and computer or tablet PC had a 60.0% higher health-related internet use score, those who owned a smartphone only had a 72.5% higher health-related internet use score, and those who had a smartphone and computer or tablet PC had an 81.5% higher health-related internet use score. The *F* statistic ($F_{17,596}$=10.30, *P*<.001) and R^2 value (0.20) indicated a good model fit to the data.

Discussion

Principal Findings

This study found that better health literacy and greater access to technological devices were associated with higher levels of health-related information-seeking behavior online. We found that the higher the health literacy level and the higher the accessibility to technological devices (eg, access to a mobile phone and computer or tablet PC at home or access to a smartphone), the more likely the respondents were to use the internet to seek health-related information. We also found differences in online health-related information-seeking behavior by sociodemographic characteristics; being female, having a graduate degree, and reporting a poor/fair health status were associated with higher usage of the internet for seeking health-related information.

It is possible that individuals with higher health literacy are more comfortable seeking out health information, are more adept at knowing what to search for and how to find it, and are more comfortable interpreting the information that they access. Our findings about access to technological devices are logical; individuals with more advanced, faster potential for online connectivity are more likely to use those devices for a variety of information-seeking purposes compared with individuals who have more limited access to technology. However, these findings signal a concerning disparity in access to information. As all types of information, including health information, are increasingly delivered online, individuals without access to efficient, effective technological devices are at risk of being left further behind.

The findings about differences in health-related information-seeking by sociodemographic characteristics indicate potential areas of inequity. For example, individuals with graduate degrees were more likely than individuals without college degrees to use the internet for health information. Again, this finding is not necessarily surprising, given already known disparities in access to and use of technology by socioeconomic status, but it could heighten disparities in access to health information [34,35]. On many of the individual items measuring online health-related information-seeking behavior, and in the multivariate model predicting ever using the internet for health information (vs never), women were more likely than men to obtain health information online. This may be related to a broader trend of gender differences in health care utilization [36]; however, this also presents an opportunity to implement more strategies to make online health information appealing and useful for men [37]. Additionally, we detected differences by age in online health information-seeking; younger adults and older adults were more likely than middle-aged adults to seek out health information online. For people of all ages who do not access health information online, it is important to ensure that comparable information is easily accessible in other forms. For example, for clinics and hospitals moving toward online-only scheduling and online communication with providers, careful thought should be given to who might potentially be left out and what alternative forms of communication, scheduling, and information delivery can and should be offered. For other sources of online health information (eg, websites, social media, hospital websites), care should be taken to ensure that people of all ages have equitable access to high-quality information, and that people who do not access such information online have equitable access to other forms of information.

Limitations

This study should be considered in light of its limitations. We relied on a cross-sectional design, which therefore limits the ability to determine causality between online behavior and health literacy. We also did not have a random sample, but rather relied on a sample of adults who attended the Minnesota State Fair and were willing to respond to a survey. Nevertheless, we collected a robust sample capable of detecting meaningful differences in online health-related information-seeking behavior. To the extent that we did not capture a fully representative sample, we are likely to understate the differences we identified. We also did not examine intersectional differences by sociodemographic characteristics (eg, the potential multiplicative impact of gender and age), which should be explored more fully in future research. Finally, although we were able to examine differences in a range of health-related information-seeking behaviors, we were not able to determine the intensity or quality of these online interactions. Not all health information delivered online is good, and more attention should be paid to ensuring high-quality content and to educating the public on how to filter good from bad information.

Practice Implications

Health information and health care are increasingly being delivered online. Prior research has shown that nearly 4 out of every 5 Americans are currently using the internet for health-related purposes [10,11]. Nearly half of the adults in our sample reported using the internet to communicate with their health care provider; more than half used it to look for information about their diet, weight, or physical activity; and more than half used the internet to look for health-related information for someone else. Clearly, the internet plays a large and growing role in how Americans manage and learn about their own health and the health of their loved ones. Thus, the internet has the potential to improve access to health information [9], to reduce barriers in communicating with health professionals [38], and to offer assistance to caregivers [39]. However, our findings also show that the internet is neither universally accessible nor universally used for obtaining health information. More attention needs to be paid to improving access to technology and to offering alternative forms of health information and communication for those without it. This might include expanding access to broadband internet and cellular connectivity in rural areas that do not have it; ensuring that public spaces such as libraries have ample access to computers where people can go online, as well as privacy protections for sensitive health information, such as cubby walls or dividers; and providing ample health information through other free sources, including libraries, clinics, and community spaces.

Finally, our study did not measure the quality of the information that individuals receive or the interactions they have. As health care, along with many other sectors of society, increasingly moves online, providers and educators are faced with the enormous task of assisting the public in filtering good information from bad and in advocating for high-quality health information online. This process might start early, in schools working with children on online literacy, but should also expand to include people of all ages, including in workplaces, senior centers, and health care settings.

Conclusion

In this study, we demonstrated differences in online health-related information-seeking behavior according to the degree of health literacy, access to technological devices, and sociodemographic characteristics. We found that individuals with better health literacy, more access to sophisticated technological devices, and more education are more likely to access health information on the internet. Although these findings may not be surprising, they should be concerning; access to information is an important predictor of behavior [40,41] and disparities in access to health information may exacerbate disparities in health and health care access. Research, policy, and programmatic attention should be paid to improving access to technology, to ensure access to alternative forms of health information for those who cannot or will not access information online, and to improve the quality of online health information.

Acknowledgments

This research project was jointly supported by the Masonic Cancer Center at the University of Minnesota, Twin Cities, and Research Fund of the Endowed Academic Chair at the University of Alabama School of Social Work awarded to HL.

Conflicts of Interest

None declared.

References

1. Meppelink CS, Smit EG, Diviani N, Van Weert JCM. Health Literacy and Online Health Information Processing: Unraveling the Underlying Mechanisms. J Health Commun 2016;21(Sup 2):109-120. [doi: 10.1080/10810730.2016.1193920] [Medline: 27668318]
2. Miller TA. Health literacy and adherence to medical treatment in chronic and acute illness: A meta-analysis. Patient Educ Couns 2016 Jul;99(7):1079-1086 [FREE Full text] [doi: 10.1016/j.pec.2016.01.020] [Medline: 26899632]
3. Matsuoka S, Tsuchihashi-Makaya M, Kayane T, Yamada M, Wakabayashi R, Kato NP, et al. Health literacy is independently associated with self-care behavior in patients with heart failure. Patient Educ Couns 2016 Jun;99(6):1026-1032. [doi: 10.1016/j.pec.2016.01.003] [Medline: 26830514]
4. Bennett IM, Chen J, Soroui JS, White S. The contribution of health literacy to disparities in self-rated health status and preventive health behaviors in older adults. Ann Fam Med 2009;7(3):204-211 [FREE Full text] [doi: 10.1370/afm.940] [Medline: 19433837]
5. Davis TC, Williams MV, Marin E, Parker RM, Glass J. Health literacy and cancer communication. CA Cancer J Clin 2002 May 01;52(3):134-149 [FREE Full text] [doi: 10.3322/canjclin.52.3.134] [Medline: 12018928]
6. White S, Chen J, Atchison R. Relationship of preventive health practices and health literacy: a national study. Am J Health Behav 2008;32(3):227-242. [doi: 10.5555/ajhb.2008.32.3.227] [Medline: 18067463]
7. Murero M, Rice RE. The Internet and Health Care: Theory, Research, and Practice. Milton Park, UK: Routledge; 2013.
8. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
9. Kalichman SC, Cain D, Cherry C, Pope H, Eaton L, Kalichman MO. Internet use among people living with HIV/AIDS: coping and health-related correlates. AIDS Patient Care STDS 2005 Jul;19(7):439-448. [doi: 10.1089/apc.2005.19.439] [Medline: 16053401]
10. Health Information National Trends Survey (HINTS). National Cancer Institute. 2014. URL: https://hints.cancer.gov/ [accessed 2020-08-15]
11. Internet/Broadband Fact Sheet. Pew Research Center. URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2020-09-01]
12. Cutrona SL, Mazor KM, Agunwamba AA, Valluri S, Wilson PM, Sadasivam RS, et al. Health Information Brokers in the General Population: An Analysis of the Health Information National Trends Survey 2013-2014. J Med Internet Res 2016 Jun 03;18(6):e123 [FREE Full text] [doi: 10.2196/jmir.5447] [Medline: 27260952]
13. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb 20;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]

14. Jiang S, Street RL. Factors Influencing Communication with Doctors via the Internet: A Cross-Sectional Analysis of 2014 HINTS Survey. Health Commun 2017 Feb;32(2):180-188. [doi: 10.1080/10410236.2015.1110867] [Medline: 27196037]
15. Rowley J, Johnson F, Sbaffi L. Gender as an influencer of online health information-seeking and evaluation behavior. J Assn Inf Sci Tec 2015 Aug 25;68(1):36-47. [doi: 10.1002/asi.23597]
16. Genuis SK. Constructing “sense” from evolving health information: A qualitative investigation of information seeking and sense making across sources. J Am Soc Inf Sci Tec 2012 Jun 29;63(8):1553-1566. [doi: 10.1002/asi.22691]
17. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
18. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
19. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
20. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The Relationship of Health Literacy With Use of Digital Technology for Health Information: Implications for Public Health Practice. J Public Health Manag Pract 2017;23(4):380-387. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]
21. Wei K, Teo H, Chan HC, Tan BCY. Conceptualizing and Testing a Social Cognitive Model of the Digital Divide. Inf Syst Res 2011 Mar;22(1):170-187. [doi: 10.1287/isre.1090.0273]
22. Zach L, Dalrymple PW, Rogers ML, Williver-Farr H. Assessing internet access and use in a medically underserved population: implications for providing enhanced health information services. Health Info Libr J 2012 Mar;29(1):61-71. [doi: 10.1111/j.1471-1842.2011.00971.x] [Medline: 22335290]
23. Benigeri M, Pluye P. Shortcomings of health information on the Internet. Health Promot Int 2003 Dec 01;18(4):381-386. [doi: 10.1093/heapro/dag409] [Medline: 14695369]
24. Friis K, Lasgaard M, Rowlands G, Osborne RH, Maindal HT. Health Literacy Mediates the Relationship Between Educational Attainment and Health Behavior: A Danish Population-Based Study. J Health Commun 2016 Sep 26;21(Sup 2):54-60. [doi: 10.1080/10810730.2016.1201175] [Medline: 27668691]
25. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010;15(Suppl 2):9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
26. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
27. Geboers B, Reijneveld SA, Jansen CJ, de Winter AF. Health Literacy Is Associated With Health Behaviors and Social Factors Among Older Adults: Results from the LifeLines Cohort Study. J Health Commun 2016 Sep 23;21(Sup 2):45-53. [doi: 10.1080/10810730.2016.1201174] [Medline: 27661472]
28. Levin-Zamir D, Baron-Epel OB, Cohen V, Elhayany A. The Association of Health Literacy with Health Behavior, Socioeconomic Indicators, and Self-Assessed Health From a National Adult Survey in Israel. J Health Commun 2016;21(Sup 2):61-68. [doi: 10.1080/10810730.2016.1207115] [Medline: 27669363]
29. Sentell TL, Tsoh JY, Davis T, Davis J, Braun KL. Low health literacy and cancer screening among Chinese Americans in California: a cross-sectional analysis. BMJ Open 2015 Jan 05;5(1):e006104 [FREE Full text] [doi: 10.1136/bmjopen-2014-006104] [Medline: 25564140]
30. Han H, Hong H, Starbird LE, Ge S, Ford AD, Renda S, et al. eHealth Literacy in People Living with HIV: Systematic Review. JMIR Public Health Surveill 2018 Sep 10;4(3):e64 [FREE Full text] [doi: 10.2196/publichealth.9687] [Medline: 30201600]
31. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
32. HINTS 4 Cycle 1 Questionnaire. National Cancer Institute (NCI), Health Information National Trends Survey (HINTS). 2014. URL: https://hints.cancer.gov/view-questions-topics/all-hints-questions.aspx [accessed 2020-07-16]
33. Behavioral Risk Factor Surveillance System Questionnaire (BRFSS) Questionnaires. Centers for Disease Control and Prevention. 2016. URL: https://www.cdc.gov/brfss/questionnaires/index.htm [accessed 2020-07-28]
34. Sedrak MS, Soto-Perez-De-Celis E, Nelson RA, Liu J, Waring ME, Lane DS, et al. Online Health Information-Seeking Among Older Women With Chronic Illness: Analysis of the Women's Health Initiative. J Med Internet Res 2020 Apr 09;22(4):e15906 [FREE Full text] [doi: 10.2196/15906] [Medline: 32271152]
35. Gordon NP, Crouch E. Digital Information Technology Use and Patient Preferences for Internet-Based Health Education Modalities: Cross-Sectional Survey Study of Middle-Aged and Older Adults With Chronic Health Conditions. JMIR Aging 2019 Apr 04;2(1):e12243 [FREE Full text] [doi: 10.2196/12243] [Medline: 31518291]
36. Green CA, Pope CR. Gender, psychosocial factors and the use of medical services: a longitudinal analysis. Soc Sci Med 1999 May;48(10):1363-1372. [doi: 10.1016/s0277-9536(98)00440-7]

37. Venkatesh V, Morris MG. Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior. MIS Quart 2000 Mar;24(1):115. [doi: 10.2307/3250981]
38. Khoong EC, Rivadeneira NA, Hiatt RA, Sarkar U. The Use of Technology for Communicating With Clinicians or Seeking Health Information in a Multilingual Urban Cohort: Cross-Sectional Survey. J Med Internet Res 2020 Apr 06;22(4):e16951 [FREE Full text] [doi: 10.2196/16951] [Medline: 32250280]
39. Milios A, McGrath P, Baillie H. A Weekly, Evidence-Based Health Letter for Caregivers (90Second Caregiver): Usability Study. JMIR Form Res 2020 Feb 12;4(2):e14496 [FREE Full text] [doi: 10.2196/14496] [Medline: 32049064]
40. Siedner MJ, Santorino D, Haberer JE, Bangsberg DR. Know your audience: predictors of success for a patient-centered texting app to augment linkage to HIV care in rural Uganda. J Med Internet Res 2015 Mar 24;17(3):e78 [FREE Full text] [doi: 10.2196/jmir.3859] [Medline: 25831269]
41. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984 Sep 04;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]

Edited by G Eysenbach; submitted 22.03.19; peer-reviewed by C Knoepke, L van Velsen, KC Wong; comments to author 03.10.19; revised version received 20.02.20; accepted 11.08.20; published 27.01.21.

Please cite as:
Lee HY, Jin SW, Henning-Smith C, Lee J, Lee J
Role of Health Literacy in Health-Related Information-Seeking Behavior Online: Cross-sectional Study
J Med Internet Res 2021;23(1):e14088
URL: http://www.jmir.org/2021/1/e14088/
doi:10.2196/14088
PMID:33502332

Original Paper

Differences in Mode Preferences, Response Rates, and Mode Effect Between Automated Email and Phone Survey Systems for Patients of Primary Care Practices: Cross-Sectional Study

Sharon Johnston[1,2], LLM, MD; William Hogg[2], BSc, MSc, MClSc, MD; Sabrina T Wong[3], RN, PhD; Fred Burge[4], MD, MSc; Sandra Peterson[3], MSc

[1]Department of Family Medicine, University of Ottawa, Ottawa, ON, Canada
[2]Department of Family Medicine, Institu du Savoir, Hôpital Montfort, University of Ottawa, Ottawa, ON, Canada
[3]Centre for Health Services and Policy Research, University of British Columbia, Vancouver, BC, Canada
[4]Department of Family Medicine, Dalhousie University, Halifax, NS, Canada

Corresponding Author:
Sharon Johnston, LLM, MD
Department of Family Medicine
University of Ottawa
43 Bruyere St
Ottawa, ON, K1N5C8
Canada
Phone: 1 613 562 6262 ext 2931
Email: sjohnston@bruyere.org

Abstract

Background: A growing number of health care practices are adopting software systems that link with their existing electronic medical records to generate outgoing phone calls, emails, or text notifications to patients for appointment reminders or practice updates. While practices are adopting this software technology for service notifications to patients, its use for collection of patient-reported measures is still nascent.

Objective: This study assessed the mode preferences, response rates, and mode effect for a practice-based automated patient survey using phone and email modalities to patients of primary care practices.

Methods: This cross-sectional study analyzed responses and respondent demographics for a short, fully automated, telephone or email patient survey sent to individuals within 72 hours of a visit to their regular primary care practice. Each survey consisted of 5 questions drawn from a larger study's patient survey that all respondents completed in the waiting room at the time of their visit. Automated patient survey responses were linked to self-reported sociodemographic information provided on the waiting room survey including age, sex, reported income, and health status.

Results: A total of 871 patients from 87 primary care practices in British Columbia, Ontario, and Nova Scotia, Canada, agreed to the automated patient survey and 470 patients (45.2%) completed all 5 questions on the automated survey. Email administration of the follow-up survey was preferred over phone-based administration, except among patients aged 75 years and older ($P<.001$). Overall, response rates for those who selected an emailed survey (369/606, 60.9%) were higher ($P<.001$) than those who selected the phone survey (101/265, 38.1%). This held true irrespective of age, sex, or chronic disease status of individuals. Response rates were also higher for email (range 57.4% [58/101] to 66.3% [108/163]) compared with phone surveys (range 36% [23/64] to 43% [10/23]) for all income groups except the lowest income quintile, which had similar response rates (email: 29/63, 46%; phone: 23/50, 46%) for phone and email modes. We observed moderate (range 64.6% [62/96] to 78.8% [282/358]) agreement between waiting room survey responses and those obtained in the follow-up automated survey. However, overall agreement in responses was poor (range 45.3% [43/95] to 46.2% [43/93]) for 2 questions relating to care coordination.

Conclusions: An automated practice-based patient experience survey achieved significantly different response rates between phone and email and increased response rates for email as income group rose. Potential mode effects for the different survey modalities may limit multimodal survey approaches. An automated minimal burden patient survey could facilitate the integration of patient-reported outcomes into care planning and service organization, supporting the move of our primary care practices toward a more responsive, patient-centered, continual learning system. However, practices must be attentive to furthering inequities in health care by underrepresenting the experience of certain groups in decision making based on the reach of different survey modes.

(J Med Internet Res 2021;23(1):e21240) doi:10.2196/21240

KEYWORDS

response rates; primary care; mixed-mode survey

Introduction

The development of an information infrastructure to support a learning health system in primary care has advanced significantly with the application of advanced analytics applied to data from electronic medical records and routinely collected administrative data [1]. However, in Canada most primary care is delivered in small community-based practices and, unlike the United Kingdom, there is no national or provincial infrastructure to measure and report patient experience data for primary care. Such data collection remains logistically challenging and relatively expensive for smaller practices [2]. While waiting room surveys often provide good response rates, they are costly, burdensome to practices, introduce a sampling bias toward older and more complex patients, and are limited to patients who physically attend a practice [1,3].

A growing number [4] of health care practices are adopting software systems [5] that link with their existing electronic medical records to generate outgoing phone calls, emails, or text notifications to patients for appointment reminders or practice updates. While practices are adopting this software technology for service notifications to patients, it is not clear whether such an approach would be acceptable to survey a practice's patients on experience or outcome measures selected by the practice to advance their quality improvement efforts. The data on response rates for electronic surveys in primary care is rudimentary compared with that for hospital surveys [4], but response rates of 20% to 30% [6,7] have been found recently for emailed surveys linked to primary care practices [6-8]. The objective of this study is to assess the mode preferences, response rates, and mode effect for a practice-based automated patient survey using phone and email modalities to patients of primary care practices.

Methods

Study Sample

This cross-sectional study analyzed mode preferences, response rates, and respondent demographics for a short, fully automated, telephone or email patient survey to consenting individuals who had recently attended their regular primary care practice. Within our larger study, Transforming Community-Based Primary Health Care Delivery through Comprehensive Performance Measurement and Reporting (TRANSFORMATION), patients from 87 primary care practices in British Columbia, Ontario, and Nova Scotia, Canada, were asked to complete a waiting room survey between 2014 and 2016. The automated patient survey system was tested on a convenience sample of those participants who consented to receiving an additional postvisit survey by email or phone. Eligible patients had to speak English or French and have a valid telephone number or email address. Patients were asked to specify their preferred contact modality, phone or email, and provide their name and contact information to an on-site research team member.

The contact information and unique identifying number for consenting patients was entered manually by survey administrators and uploaded to a cloud-based server via a software console. Upon receipt of the information, the administering information technology company collaborator, Cliniconex [9], programmed the appropriate survey mode and language (English or French) and randomly assigned the order of 5 survey questions. Once the survey was administered by Cliniconex, all contact information was deleted, and only the unique identifying number was retained on the server.

Survey Administration

Participants received an automated phone or email survey within 72 hours of visiting the practice. A phone survey response was recorded as completed only if the patient could be reached at the phone number on file, accepted the call, and completed all 5 survey questions. The phone survey was initially attempted twice, and then registered as incomplete if no answer was received. Partway through the study, the number of attempts was increased (4 call attempts) to facilitate higher response rates. For those who chose an email survey mode, an email was sent once containing the introduction and a web link to the survey. An email survey was recorded as complete if all 5 survey questions were answered.

Each survey consisted of 5 questions drawn from the TRANSFORMATION study's waiting room patient survey [10]. The questions were selected to relate to patients' experience with primary care and/or their practice. Two question prompts were modified from their original form in the paper waiting room survey to reflect the timing of the survey administration. When administered in the waiting room, questions 1 and 2 were prompted with "After seeing the family doctor or nurse today..."; on the automated patient survey, patients were prompted with "At your last visit with your family doctor or nurse practitioner...." See Multimedia Appendix 1 for the wording of the survey questions in the paper waiting room survey and the postvisit automated survey. Phone survey responses were stored in a secure password-protected site on a secure server. Email responses were sent directly to a hospital-based server and managed using Research Electronic Data Capture tools [11].

The unique identifying numbers were used to link automated patient survey responses to self-reported sociodemographic information on the paper waiting room survey, completed during the participant's visit to their practice, including age, sex, reported income, and health status.

Data Analysis

To detect any response bias inherent in using an automated email or phone survey system, we used Pearson chi-square tests to compare the sociodemographic profile of those who completed the automated patient survey (responders) with those who did not complete the automated patient survey (nonresponders). The comparison group of nonresponders contained those who either participated in the paper waiting room survey but refused the automated survey or agreed to the automated survey but did not complete all 5 questions. We also conducted Wilcoxon rank-sum tests on the paper waiting room survey responses, comparing differences in mean responses between those who completed the automated patient survey and those who did not. We conducted chi-square tests to compare automated patient survey mode preference (email or phone) and response rates both across and between patient sociodemographics. A Cochran-Armitage test for trends was also used to examine variation in mode preference by age and income. All analyses were performed using SAS software version 9.4 (SAS Institute Inc).

The primary outcome measure response rate was pooled across all practices as we were interested in differences across dependent variables of age and attributed socioeconomic status rather than regional variations.

To identify a potential mode effect, secondary analyses explored responses for each question across the 3 survey modes, email and phone (automated patient survey) and paper (waiting room). Test-retest analysis was undertaken, comparing each patient's responses from the waiting room survey to their responses to the corresponding automated survey question. The percentage of concordant and discordant responses were determined by comparing waiting room derived responses with those from subsequent survey data. Weighted kappas were calculated to compare this concordance in survey responses by survey mode. Mean responses were also compared (using Wilcoxon signed-rank test) across responses to the corresponding questions from the paper waiting room survey and the automated patient survey (total and by mode).

This study was approved by the behavioral research ethics boards at Fraser Health (RHREB 2015-017), University of British Columbia (H13-01237), Ottawa Health Science Network (20140485-01H), Bruyère Continuing Care (M16-14-029), and the Nova Scotia Health Authority (CDHA-RS/2015-150).

Results

Response Bias

Of those who agreed to the automated patient survey, 69.6% (606/871) of participants chose to receive the survey by email compared with telephone. This group represented 45.15% (871/1929) of the participants who initially consented to completing a paper waiting room survey (Table 1). Of those who agreed to the survey, 55.6% (484/871) responded and 97.1% (470/484) completed all 5 questions (24.36% [470/1929] of those who completed the paper waiting room survey and 54.0% [470/871] of those who agreed to the automated patient survey). Respondents to the automated patient survey tended to be older, were more likely to be women, had higher income, and reported a larger number of chronic conditions than those not completing the survey. There was no significant difference in paper waiting room survey responses between those who completed the automated patient survey and those who did not (Table 2).

Table 1. Comparison of those who completed the automated patient survey to those who did not[a].

Characteristics	Total	Completed automated patient survey			Did not complete automated patient survey[b]
		Phone	Email	Total[c]	
Consented to automated patient survey, n (%)	871 (45.2)	N/A[d]	N/A	N/A	N/A
Automated patient survey response rate, n (%)	484 (55.6)	N/A	N/A	N/A	N/A
Automated patient survey completion rate (all 5 questions), n (%)	470 (54.0)	N/A	N/A	N/A	N/A
Overall, n (%)	1929	97 (5.0)	361 (18.7)	458 (23.7)	1471 (76.3)
Age group, n (%)					
18-24	99	0 (0)	11 (3.0)	11 (2.4)	88 (6.0)
25-64	1251	50 (51.5)	263 (72.9)	313 (68.3)	938 (64.2)
65-74	377	35 (36.1)	62 (17.2)	97 (21.2)	280 (19.2)
75+	191	12 (12.4)	25 (6.9)	37 (8.1)	154 (10.5)
χ^2 *P* value[e]	N/A	<.001	.006	.006	N/A
Cochran-Armitage 2-sided *P* value[f]	N/A	<.001	.14	.86	N/A
Sex, n (%)					
Male	634	28 (29.2)	97 (27.0)	125 (27.5)	509 (35.3)
Female	1263	68 (70.8)	262 (73.0)	330 (72.5)	933 (64.7)
χ^2 *P* value[e]	N/A	.22	.003	.002	N/A
Income ($), n (%)					
<20,000	266	23 (25.6)	29 (8.5)	52 (12.0)	214 (17.0)
20,000-40,000	367	23 (25.6)	58 (17.0)	81 (18.8)	286 (22.7)
40,000-60,000	340	18 (20.0)	74 (21.6)	92 (21.3)	248 (19.7)
60,000-100,000	435	16 (17.8)	108 (31.6)	124 (28.7)	311 (24.7)
>100,000	283	10 (11.1)	73 (21.3)	83 (19.2)	200 (15.9)
χ^2 *P* value	N/A	.15	<.001	.02	N/A
Cochran-Armitage 2-sided *P* value[f]	N/A	.01	<.001	<.001	N/A
Chronic conditions, n (%)					
0-1	637	12 (12.4)	115 (31.9)	127 (27.8)	510 (36.1)
2	291	14 (14.4)	48 (13.3)	62 (13.6)	229 (16.2)
3+	943	71 (73.2)	197 (54.7)	268 (58.6)	675 (47.7)
χ^2 *P* value	N/A	<.001	.06	<.001	N/A

[a] Total counts within categories vary due to missing data.

[b] Includes those who did not consent to having their automated patient survey linked to their waiting room survey responses.

[c] Excludes those who did not consent to having their automated patient survey linked to their waiting room survey responses.

[d] N/A: not applicable.

[e] χ^2 *P* value is comparing the distribution of the subgroup (eg, age, gender) between completed and not completed, for each mode and overall.

[f] Cochrane-Armitage *P* value for the presence of linear trend in proportions of completed and not completed across ordinal subcategories, for each mode and overall.

Table 2. Comparison of waiting room survey responses between those who completed the automated patient survey to those who did not[a].

Waiting room survey questions	Response for those who also completed automated patient survey[b], mean (SD)	Response for those who did not complete automated patient survey[c], mean (SD)	*P* value[d]
Given enough time	4.67 (0.60)	4.63 (0.75)	.88
Explained tests and treatments	4.70 (0.61)	4.66 (0.71)	.85
Told about potential side effects of medications	2.64 (0.60)	2.69 (0.55)	.30
Times when provider didn't have access to recent tests or exam results	1.26 (0.51)	1.25 (0.52)	.63
Times when provider didn't know about changes in treatment plan that another person recommended	1.23 (0.47)	1.20 (0.44)	.37

[a] Total counts within categories vary due to missing data.

[b] Excludes those who did not consent to having their automated patient survey linked to their waiting room survey responses.

[c] Includes those who did not consent to having their automated patient survey linked to their waiting room survey responses.

[d] Wilcoxon rank-sum test *P* value is comparing paper waiting room survey responses between completed and not completed.

Response Rates

In this sample, email administration of the follow-up survey was preferred over phone-based administration, except among patients aged 75 years and older (Table 3). Among those who answered the automated patient survey, 97.1% (470/484) completed of all 5 questions. Thus, response rates include only those who answered all 5 questions. Overall, response rates for those who selected an emailed survey (369/606, 60.9%) were higher than those who received the phone survey (101/265, 38.1%). This held true irrespective of the age, sex, or chronic disease status of individuals. Response rates were also higher for email compared with phone surveys for all income groups except the lowest income quintile, which had similar response rates for phone and email modes. There was variation in response rates within email mode, with higher responses among more affluent individuals.

Table 3. Mode preference and response rates by subgroups.

Characteristic	Mode preference (n=871)				Completed automated patient survey (n=470)			
	Total, n (%)	Phone, n (%)	Email, n (%)	χ^2 *P* value[a]	Total, n (%)	Phone, n (%)	Email, n (%)	χ^2 *P* value[b]
Age group, years	N/A[c]	N/A	N/A	<.001[d]	N/A	N/A	N/A	N/A
18-24	29 (3.4)	s[e]	26 (89.7)	N/A	11 (37.9)	0 (0)	11 (42.3)	.15
25-64	553 (65.0)	133 (24.1)	420 (75.9)	N/A	313 (56.6)	50 (37.6)	263 (62.6)	<.001
65-74	185 (21.7)	78 (42.2)	107 (57.8)	N/A	97 (52.4)	35 (44.9)	62 (57.9)	.08
75+	84 (9.9)	45 (53.6)	39 (46.4)	N/A	37 (44.0)	12 (26.7)	25,871 (64.1)	<.001
χ^2 *P* value[f]	N/A	N/A	N/A	N/A	.045	.11	.18	N/A
Cochran-Armitage 2-sided *P* value[g]	N/A	N/A	N/A	N/A	.16	.64	.60	N/A
Sex	N/A	N/A	N/A	<.001	N/A	N/A	N/A	N/A
Male	262 (31.1)	101 (38.5)	161 (61.5)	N/A	125 (47.7)	28 (27.7)	97 (60.2)	<.001
Female	581 (68.9)	154 (26.5)	427 (73.5)	N/A	330 (56.8)	68 (44.2)	262 (61.4)	<.001
χ^2 *P* value	N/A	N/A	N/A	N/A	.01	.008	.81	N/A
Income ($)	N/A	N/A	N/A	<.001[d]	N/A	N/A	N/A	N/A
<20,000	113 (14.5)	50 (44.2)	63 (55.8)	N/A	52 (46.0)	23 (46.0)	29 (46.0)	>.99
20,000-40,000	165 (21.1)	64 (38.8)	101 (61.2)	N/A	81 (49.1)	23 (35.9)	58 (57.4)	.007
40,000-60,000	162 (20.7)	47 (29.0)	115 (71.0)	N/A	92 (56.8)	18 (38.3)	74 (64.3)	.002
60,000-100,000	206 (26.3)	43 (20.9)	163 (79.1)	N/A	124 (60.2)	16 (37.2)	108 (66.3)	<.001
100,000+	136 (17.4)	23 (16.9)	113 (83.1)	N/A	83 (61.0)	10 (43.5)	73 (64.6)	.06
χ^2 *P* value	N/A	N/A	N/A	N/A	.03	.83	.049	N/A
Cochran-Armitage 2-sided *P* value	N/A	N/A	N/A	N/A	.002	.73	.01	N/A
Income ($)	N/A	N/A	N/A	<.001	N/A	N/A	N/A	N/A
<20,000	113 (14.5)	50 (44.2)	63 (55.8)	N/A	52 (46.0)	23 (46.0)	29 (46.0)	NS >.99
20,000+	669 (85.5)	177 (26.5)	492 (73.5)	N/A	380 (56.8)	67 (37.9)	313 (63.6)	<.001
χ^2 *P* value	N/A	N/A	N/A	N/A	.03	.30	.007	N/A
# Chronic conditions	N/A	N/A	N/A	<.001	N/A	N/A	N/A	N/A
0-1	242 (28.7)	43 (17.8)	199 (82.2)	N/A	127 (52.5)	12 (27.9)	115 (57.8)	<.001
2	126 (14.9)	42 (33.3)	84 (66.7)	N/A	62 (49.2)	14 (33.3)	48 (57.1)	.01
3+	475 (56.4)	170 (35.8)	305 (64.2)	N/A	268 (56.4)	71 (41.8)	197 (64.6)	<.001
χ^2 *P* value	N/A	N/A	N/A	N/A	.29	.20	.22	N/A

[a] Comparing percentage distribution of mode preference across subgroups.

[b] Comparing percentage completed across subgroups.

[c] N/A: not applicable.

[d] Cochran-Armitage test for trend also gives *P*<.001.

[e] Indicates suppressed due to cell count less than 5.

[f] Comparing completion rates between subgroups (overall or by mode).

[g] Cochran-Armitage conducted to test for trends.

Mode Effect

We observed moderate agreement between waiting room survey responses and those obtained in the follow-up automated survey (see Multimedia Appendix 2). However, overall agreement in responses was poor for 2 questions relating to care coordination.

Among phone respondents, agreement in responses was generally poor, and phone responders were particularly critical with respect to care coordination (Table 4). Agreement between waiting room responses and subsequent email survey regarding interpersonal aspects of care was moderate and poor for items relating to care coordination.

Table 4. Comparison of responses to paper waiting room surveys and automated surveys among those who completed the automated patient survey.

Description	Waiting room response, mean (SD)	Automated patient survey response—by mode			
		Phone, mean[a] (SD)	*P* value[b,c]	Email, mean[d] (SD)	*P* value[c,e]
Given enough time (range 1-5[f])	4.67 (0.60)	4.66 (0.72)	.15	4.67 (0.65)	.27
Explained tests and treatments (range 1-5[f])	4.70 (0.61)	4.42 (0.97)	.04	4.57 (0.79)	<.001
Told about potential side effects of medications (range 1-3[g])	2.64 (0.60)	2.63 (0.70)	>.99	2.68 (0.68)	.28
Times when provider didn't have access to recent tests or exam results (range 1-3[h])	1.26 (0.51)	1.97 (0.90)	<.001	1.65 (0.79)	<.001
Times when provider didn't know about changes in treatment plan that another person recommended (range 1-3[h])	1.23 (0.47)	1.94 (0.88)	<.001	1.52 (0.71)	<.001

[a] The final N available for analysis for each question varies slightly due to nonresponse (or not applicable choice). Mean responses are calculated only for those who answered both versions of the questions. Ns for each question for those who completed the automated patient survey by phone are as follows: Q1: 96; Q2: 96; Q3: 92; Q4: 93.

[b] Comparing phone automated patient survey response to waiting room response.

[c] Wilcoxon signed-rank test.

[d] The final N available for analysis for each question varies slightly due to nonresponse (or not applicable choice). Mean responses and paired mean differences are calculated only for those who answered both versions of the questions. Ns for each question for those who completed the automated patient survey by email are as follows: Q1: 358; Q2: 357; Q3: 290; Q4: 336; Q5: 360.

[e] Comparing email automated patient survey response to waiting room response.

[f] 1=very poor; 5=very good.

[g] 1=no; 3=yes, often, or always.

[h] 1=never or rarely; 3=often or very often.

Discussion

Principal Findings

We successfully deployed an automated multimodal practice-based patient survey in 87 primary care practices. Overall, patient preference for the email survey mode was demonstrated; however, this was modified by age group and socioeconomic status. Indeed, completion rates for email were higher than most health care automated surveys [8] versus comparable response rates in the total sample [6]. However, it is unclear whether the lower consent rate (45.2%) from the total patient sample reflects lack of acceptability of an automated low-burden survey or survey fatigue among participants who had already completed a long waiting room survey. Despite this, the relatively high completion rate to the short email survey suggests this is a feasible and acceptable approach to collect patient reported data.

Our results show that the lowest income group had the lowest preference for the email mode and lowest response rate for the email survey while having the highest response rates for the phone survey. Our finding of the email responders being more likely to be female and of higher income echoes the pattern of a recent practice-based single-site email survey in Ontario [6]. A move to use email surveys to collect patient experience data would need to carefully monitor underrepresentation by lowest income groups to not exacerbate inequities in health care. The survey software, as it is used currently for appointment reminders, is usually deployed after linking with the electronic medical record to use patient contact information, so it is possible for automated surveys to track information such as approximate income based on postal code and oversample a population found to be underrepresented in responses.

Opportunities to match surveys to reported language preferences and the capacity to reach people by phone or email who do not frequently attend a practice or have a stable home address raises the potential for an automated survey to be particularly valuable in understanding the experience of some of the most vulnerable members of a practice population. However, there are still inequities in access to the internet, with lower income individuals and people living in rural areas having lower access [12-14]. Text messaging might be preferable to phone for some patients and increase the reach across sociodemographic groups.

The low concordance rate of responses on questions of care coordination between paper and automated survey, especially the phone survey, raises important questions about a mode effect and/or the role of true anonymity in responding to questions about one's health care provider or practice in a waiting room compared with online or automated phone response. It is also possible that the paper survey questions on care coordination sensitized participants to the issue, and after their visit, they were more aware of breakdowns in optimal care, accounting for their more negative responses with the automated survey following their practice visit. Additionally, the care coordination questions had negative phrasing, which may have been more confusing for phone responders.

Cost-effectiveness was not the focus of this study. However, at two-thirds the completion rate compared with email, a phone survey would cost one and a half times as much. The cost of deploying a tailored automated patient outreach message and linked survey from the software company we collaborated with includes a 1-time practice start-up fee of $500 CAN (US $390) and an annual per-provider fee of $600 CAN (US $468). For an average practice of 4 providers and 5200 patients, an email survey would cost about 25 cents (US 20 cents) if each patient were sent a message and survey twice per year or less than 15 cents (US 12 cents) if most patients were sent a survey 4 times per year. Higher response rates make the approach more cost-effective for the email mode since automated systems frequently charge per survey sent. For quality improvement data collection, practices would not need to seek prior consent to contact patients. However, efforts to enhance patient buy-in and achieve higher response rates would be key to the cost-effectiveness of this approach. As practices seek better ways to engage patients and collect patient-reported experience measures and patient-reported outcome measures, it is essential to be sensitive to the response burden on patients and promote a culture in which patients understand the purpose of surveys and feel their insights and time are valued [15]. This may help build a partnership with patients in practice-based surveying as a way to give patients more influence in the system and their care.

The capacity of this proposed system to link collection of patient-reported measures with clinical services, such as appointment reminders or preventive care reminders, could improve the response rates received on general surveys of patient experiences, improving quality and reducing costs [2]. Such an approach would have the benefit of being able to deploy surveys to all patients or ones with prespecified criteria (eg, people who just attended the practice, have not attended in over a year, have a recent hospital discharge). Such a survey could be linked with data automatically extracted from electronic medical records or a registry developed by providers, offering an even greater opportunity to understand patient experiences and outcomes. Additionally, an automated system can spread the burden of response across a wide and/or randomly selected segment of a practice's patient population, asking different questions to different patients on an ongoing or rolling basis, enhancing reach and reducing cost compared with traditional waiting room surveys.

Increasingly, electronic medical records are being used to collect patient-reported outcome measures that are inputted directly into the patients' chart. This approach offers the benefit of supporting a patients' immediate care. However, this approach creates a burden for the provider or practice to review data automatically put into a patient chart in a timely manner. Keeping a patient automated survey function distinct from clinical care may be attractive to providers and practices who need to manage their workflow and feel overburdened with data and data requests already.

As a survey method, an automated patient survey offers some attractive features. Response rates and sample bias can be easily calculated for parameters such as age, gender or income as estimated by postal code without adding to patient burden of filling this information in. Based on continually updated information on filled surveys, ongoing distribution (sampling) parameters can be set to minimize or account for any bias that may arise. Automated surveys can be deployed at regular intervals determined by the practice and would not burden practice staff, providers, or even patients during a visit, thus avoiding interruptions or additional work.

As more practices are collecting email addresses from their patients and patients expect email communication options, an automated patient engagement system with an embedded survey is feasible. Practices already using this or a similar technology to serve patients through outreach reminders may be more willing to participate in data collection initiatives that use this same infrastructure for quality improvement or research.

Limitations

There are some limitations to consider in interpreting the findings of this study. Initial recruitment into the TRANSFORMATION study was through a convenience sample of patients from primary care practices across British Columbia, Ontario, and Nova Scotia. As such, patients who were recruited into the study may not be representative of patients generally across Canada, potentially limiting generalizability. Additionally, potential for selection bias is further compounded by relatively low overall response rates by participants of the automated patient survey, who were recruited from the initial convenience sample of patients enrolled into the larger study.

Conclusions

An automated practice-based patient experience survey achieved significantly different response rates between phone and email and increased response rates for email as income group rose. The higher response rates of the email surveys make a phone approach less cost-effective. However, care must be paid to furthering inequities in health care by underrepresenting the experience of certain groups in decision making. Further, potential mode effects for the different survey modalities may limit multimodal survey approaches.

An automated communication system will become even more valuable as the stock of high-quality and validated instruments to measure patient-reported outcomes grows over the next decade [16]. An automated system that enables targeted outreach surveys with minimal burden on patients and providers could facilitate the integration of patient-reported outcomes into care planning and service organization, supporting the move of our primary care practices toward a more responsive, patient-centered, continual learning system.

Acknowledgments

The authors thank Stephanie Blackman, Martha Foley, and Jonathan Beaumier for their help in writing the manuscript.

Authors' Contributions

SJ conceived the study, oversaw the implementation and analyses, and wrote the manuscript. WH helped conceive the study, oversaw the implementation, and contributed to the writing of the manuscript. SW contributed to the analyses and reviewed and approved the final manuscript. FB contributed to the analyses and reviewed and approved the final manuscript. SP led the analyses and reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions and response options for the automated patient surveys and paper waiting room surveys.
[DOCX File , 16 KB - jmir_v23i1e21240_app1.docx]

Multimedia Appendix 2
Concordance of automated patient survey responses compared with paper waiting room survey responses.
[DOCX File , 17 KB - jmir_v23i1e21240_app2.docx]

References

1. Green ME, Hogg W, Savage C, Johnston S, Russell G, Jaakkimainen RL, et al. Assessing methods for measurement of clinical outcomes and quality of care in primary care practices. BMC Health Serv Res 2012;12:214 [FREE Full text] [doi: 10.1186/1472-6963-12-214] [Medline: 22824551]
2. Peters M, Crocker H, Jenkinson C, Doll H, Fitzpatrick R. The routine collection of patient-reported outcome measures (PROMs) for long-term conditions in primary care: a cohort survey. BMJ Open 2014 Feb 21;4(2):e003968 [FREE Full text] [doi: 10.1136/bmjopen-2013-003968] [Medline: 24561495]
3. Hogg W, Johnston S, Russell G, Dahrouge S, Gyorfi-Dyke E, Kristjanssonn E. Conducting waiting room surveys in practice-based primary care research: a user's guide. Can Fam Physician 2010 Dec;56(12):1375-1376 [FREE Full text] [Medline: 21156900]
4. Khanbhai M, Flott K, Darzi A, Mayer E. Evaluating digital maturity and patient acceptability of real-time patient experience feedback systems: systematic review. J Med Internet Res 2019 Jan 14;21(1):e9076. [doi: 10.2196/jmir.9076] [Medline: 31344680]
5. McLean SM, Booth A, Gee M, Salway S, Cobb M, Bhanbhro S, et al. Appointment reminder systems are effective but not optimal: results of a systematic review and evidence synthesis employing realist principles. Patient Prefer Adherence 2016;10:479-499 [FREE Full text] [doi: 10.2147/PPA.S93046] [Medline: 27110102]
6. Slater M, Kiran T. Measuring the patient experience in primary care: comparing e-mail and waiting room survey delivery in a family health team. Can Fam Physician 2016 Dec;62(12):e740-e748 [FREE Full text] [Medline: 27965350]
7. Poppelwell E, Esplin J, Doust E, Swansson J. Evaluation of the primary care patient experience survey tool. New Zealand Ministry of Health. 2018 Apr 18. URL: https://www.hqsc.govt.nz/assets/Health-Quality-Evaluation/PES/MoH-PES-report-18April2018_2.pdf [accessed 2020-12-02]
8. Falconi M, Johnston S, Hogg W. A scoping review to explore the suitability of interactive voice response to conduct automated performance measurement of the patient’s experience in primary care. Prim Health Care Res Dev 2015 Aug 5;17(03):209-225. [doi: 10.1017/s1463423615000407] [Medline: 26242171]
9. Cliniconex. URL: http://cliniconex.com/ [accessed 2020-12-02]
10. Wong S, Burge F, Johnston S, Hogg W, Haggerty J. The TRANSFORMATION primary health care patient experiences survey in French and English: a technical report. UBC Centre for Health Services and Policy Research. 2019. URL: http://chspr.sites.olt.ubc.ca/files/2019/05/TRANSFORMATION-Pt-Exp-Survey-2019.pdf [accessed 2020-12-02]
11. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
12. CIRA: Canadian Internet Registration Authority. URL: https://cira.ca/factbook/2014/the-canadian-internet.html [accessed 2020-12-02]

13. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
14. Allen M. Consumption of culture by older Canadians on the internet. Statistics Canada. 2013. URL: http://www.statcan.gc.ca/pub/75-006-x/2013001/article/11768-eng.htm [accessed 2020-12-02]
15. Primary care patient experience survey: support guide. Health Quality Ontario. 2015 Apr. URL: http://www.hqontario.ca/Portals/0/documents/qi/primary-care/primary-care-patient-experience-survey-support-guide-en.pdf [accessed 2020-12-02]
16. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]

Abbreviations

TRANSFORMATION: Transforming Community-Based Primary Health Care Delivery Through Comprehensive Performance Measurement and Reporting

Edited by G Eysenbach; submitted 08.06.20; peer-reviewed by H van Marwijk, T Ungar; comments to author 25.08.20; revised version received 23.09.20; accepted 28.10.20; published 11.01.21.

Please cite as:
Johnston S, Hogg W, Wong ST, Burge F, Peterson S
Differences in Mode Preferences, Response Rates, and Mode Effect Between Automated Email and Phone Survey Systems for Patients of Primary Care Practices: Cross-Sectional Study
J Med Internet Res 2021;23(1):e21240
URL: https://www.jmir.org/2021/1/e21240
doi:10.2196/21240
PMID:33427675

Original Paper

Risk Stratification for Early Detection of Diabetes and Hypertension in Resource-Limited Settings: Machine Learning Analysis

Justin J Boutilier[1], PhD; Timothy C Y Chan[2], PhD; Manish Ranjan[3], MBA; Sarang Deo[4], PhD

[1]Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States
[2]Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada
[3]NanoHealth, NanoCare Health Services, Hyderabad, India
[4]Max Institute of Healthcare Management, Indian School of Business, Hyderabad, India

Corresponding Author:
Justin J Boutilier, PhD
Department of Industrial and Systems Engineering
University of Wisconsin-Madison
1513 University Avenue
Madison, WI, 53706
United States
Phone: 1 6082630350
Email: jboutilier@wisc.edu

Abstract

Background: The impending scale up of noncommunicable disease screening programs in low- and middle-income countries coupled with limited health resources require that such programs be as accurate as possible at identifying patients at high risk.

Objective: The aim of this study was to develop machine learning–based risk stratification algorithms for diabetes and hypertension that are tailored for the at-risk population served by community-based screening programs in low-resource settings.

Methods: We trained and tested our models by using data from 2278 patients collected by community health workers through door-to-door and camp-based screenings in the urban slums of Hyderabad, India between July 14, 2015 and April 21, 2018. We determined the best models for predicting short-term (2-month) risk of diabetes and hypertension (a model for diabetes and a model for hypertension) and compared these models to previously developed risk scores from the United States and the United Kingdom by using prediction accuracy as characterized by the area under the receiver operating characteristic curve (AUC) and the number of false negatives.

Results: We found that models based on random forest had the highest prediction accuracy for both diseases and were able to outperform the US and UK risk scores in terms of AUC by 35.5% for diabetes (improvement of 0.239 from 0.671 to 0.910) and 13.5% for hypertension (improvement of 0.094 from 0.698 to 0.792). For a fixed screening specificity of 0.9, the random forest model was able to reduce the expected number of false negatives by 620 patients per 1000 screenings for diabetes and 220 patients per 1000 screenings for hypertension. This improvement reduces the cost of incorrect risk stratification by US $1.99 (or 35%) per screening for diabetes and US $1.60 (or 21%) per screening for hypertension.

Conclusions: In the next decade, health systems in many countries are planning to spend significant resources on noncommunicable disease screening programs and our study demonstrates that machine learning models can be leveraged by these programs to effectively utilize limited resources by improving risk stratification.

(J Med Internet Res 2021;23(1):e20123) doi:10.2196/20123

KEYWORDS

machine learning; diabetes; hypertension; screening; global health

Introduction

Noncommunicable diseases, including diabetes, hypertension, and cardiovascular disease, are a global health priority [1]. Noncommunicable diseases disproportionally affect low- and middle-income countries, wherein more than 75% of all noncommunicable disease deaths (~31 million per year) occur, including over 16 million annual deaths in adults between the ages of 30 years and 69 years [1]. India faces the largest burden of noncommunicable diseases in the world with an estimated

73 million reported with diabetes and over 400 million people reported with hypertension [2,3]. Moreover, an estimated 58% of the patients with diabetes and 60%-75% of the patients with hypertension in India are undiagnosed, thereby creating a population health crisis [4]. Early detection via screening and subsequent treatment initiation can significantly reduce the burden of both diabetes and hypertension [5,6]. However, health systems in many low- and middle-income countries are already overburdened with an unfinished agenda on infectious diseases [7] and do not have enough capacity to conduct national-level noncommunicable disease screening programs [8].

Community-based screening programs can be leveraged to augment the capacity of the existing health systems by using community health workers (with limited training to conduct diabetes and hypertension screening) [9]. Owing to the large number of undiagnosed patients and the lack of awareness of noncommunicable diseases, community health workers typically conduct door-to-door and camp-based (ie, a tent staffed with community health workers) screenings to identify patients with undiagnosed diabetes and hypertension and subsequently refer them to a physician for assessment.

Community-based screening programs in low- and middle-income countries typically employ risk stratification methods that have been developed in high-income countries, leading to 3 key limitations [10,11]. First, at-risk populations in low- and middle-income countries differ significantly in social, lifestyle, and genetic aspects, thereby limiting the validity of models from high-income countries [12,13]. Second, a community-based approach severely limits the amount and complexity of data that can be collected by community health workers. Consequently, many models from high-income countries, which rely on advanced data (eg, triglyceride levels for diabetes [14]), are not applicable. Third, models from high-income countries are often calibrated to estimate the long-term risk of developing the disease [11,15] (eg, 2-10 years) rather than identifying the short-term risk of developing the disease. Owing to these limitations, application of approaches from high-income countries to community-based screening programs in low- and middle-income countries can result in poor risk stratification accuracy, reduced screening program yield, and increased cost per case identified [16].

In this study, we developed new risk stratification algorithms that are tailored for community-based screening programs in low- and middle-income countries with limited screening data. In particular, we used data collected by community health workers in Hyderabad, India and developed risk stratification models to estimate the short-term (2-month) risk for both diabetes mellitus and primary hypertension. We compared our results with several approaches from the literature, including previously developed risk scores from the United States and the United Kingdom. We also analyzed the trade-off between model accuracy and data availability by quantifying the incremental value of each data type collected during screening. Lastly, we quantified the expected reduction in the number of patients incorrectly stratified and the expected cost of incorrect risk stratification per patient.

Methods

Study Setting

Our study was based in Hyderabad, the capital of the state of Telangana and the fourth largest city in India with a population of 7 million [17]. Hyderabad has more than 1.7 million people living in 1400 urban slums [18]. Our catchment area included 52 urban slums and the surrounding communities. This population comprises individuals working as drivers, daily wage earners, domestic helpers, vendors, and self-employed professionals in the unorganized sector of the economy. The average family income of these residents ranges between INR 15,000 and INR 30,000 per month, which is equivalent to US $200-US $400, while the median income in Hyderabad is INR 25,000 (US $1=INR 75) per month [19].

Data Collection

We obtained retrospective data from a social enterprise based in Hyderabad that provides screening and management of diabetes and hypertension for low-income households. These data were collected through door-to-door and camp-based screenings conducted in the urban slums of Hyderabad between July 14, 2015 and April 21, 2018 by community health workers. The door-to-door screenings were conducted in low-income areas throughout Hyderabad. In our context, "camp-based" refers to individuals that were screened at a "screening camp"—a tent staffed with community health workers and setup in a low-income area. These "screening camps" were conducted in the same locations that community health workers conducted door-to-door screenings and allowed individuals to present themselves for screening.

Each community health worker was equipped with a "Doc-in-the-Bag" kit that included a weighing scale, measuring tape, blood glucose monitor, and blood pressure/heart rate cuffs (see Figure S1 of Multimedia Appendix 1 for photographs of the kit). A mobile tablet was used to record patients' responses to the questionnaire about family history, lifestyle, demographics, symptoms of common ailments, and to record certain anthropometric measurements and vitals (see the Data Description or Screening Questionnaire sections in Multimedia Appendix 1 for more details).

We included all individuals who visited a physician within 2 months following the screening to be assessed for a diagnosis of diabetes and hypertension. Hypertension was diagnosed based on 2 physician visits by using the JNC-VII (Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure) definition of hypertension [2]. Diabetes was diagnosed based on one of the following criteria: glycated hemoglobin (HbA_{1c}) higher than 6.5% (48 mmol/mol) or fasting blood glucose level higher than 126 mg/dL (7 mmol/L) [10].

Data Analysis

We developed separate models to estimate the risks for diabetes and hypertension. In each model, the outcome variable (the target) was a binary variable, indicating that a physician made a positive disease diagnosis and the predictor variables (features) were obtained from the screening data. Our primary analysis focused on determining the best models for predicting diabetes and hypertension risk and compared these models to previously developed risk scores from the United States and the United Kingdom. We compared the performance of 5 commonly used risk prediction models: decision trees, regularized logistic regression, K-nearest neighbors, random forest, and AdaBoost decision trees. We compared the performance of these models with several baseline approaches from the United States and the United Kingdom (see Table S2 of Multimedia Appendix 1). For each baseline, we calculated 3 variants (where applicable): (1) the original version based on a regression model with parameters derived from the original dataset, (2) an approximate score-based version that is intended for ease of computation by patients and providers, and (3) a version of the original regression model retrained using our Hyderabad data. In total, we considered 6 baseline models for diabetes (2 original, 2 score-based, and 2 retrained) and 2 baseline models for hypertension (1 original and 1 retrained). The hypertension model that we considered did not have an approximate score-based version (see the Baseline Approaches section in Multimedia Appendix 1 for more details).

In line with common practice, we used disjoint training (train the model), validation (tune the hyperparameters), and testing sets (test the model). We used a 10-fold cross-validation procedure to partition our data into training and testing sets. We then used 3-fold cross validation on the training set to choose our hyperparameters (see the Hyperparameter Tuning section in Multimedia Appendix 1 for more details). Once the final hyperparameters were selected, we applied the final model to the test set (that was not used as part of the model selection or fitting process in any way) to estimate generalization. We repeated the entire process 25 times to obtain more robust estimates and error bars.

For each model, we varied the discriminant threshold applied to the test set, calculated the resulting true and false positive rates, plotted them in the form of a receiver operating characteristic curve, and calculated the area under the receiver operating characteristic curve (AUC), which we used as a metric to compare different models. In total, we generated 250 test sets (25 repetitions * 10-fold) receiver operating characteristic curves for each model.

As a secondary analysis, we compared the performance of our risk stratification models, each trained with 5 different feature sets, where each set incrementally adds measurements obtained using an additional device: (1) only the questionnaire (no device measurements), (2) questionnaire and weight (weighing scale), (3) questionnaire, weight, height, and waist circumference (weighing scale and tape measure), (4) questionnaire, weight, height, waist circumference, blood pressure, and heart rate (weighing scale, tape measure, and blood pressure/heart rate cuffs), and (5) questionnaire, weight, height, waist circumference, blood pressure, heart rate, and blood glucose (weighing scale, tape measure, blood pressure/heart rate cuffs, and glucometer).

Finally, we performed a cost analysis to estimate the expected cost of incorrect risk stratification per screening for both diabetes and hypertension. We relied on previous research to estimate the cost of false positives, the cost of false negatives, and disease prevalence in India for both diabetes and hypertension. We also conducted a sensitivity analysis on each component used to estimate the expected cost of incorrect risk stratification per screening (see the Cost Analysis section of Multimedia Appendix 1 for more details).

Statistical Analysis

For all model comparisons, we conducted a 2-sided Wilcoxon signed rank test [20] (with a significance level of .05) to check whether the medians of the AUC distributions (or cost distributions) of the 2 models were different from each other. All models and statistical tests were implemented using Python 3.5, SciPy package, and the Scikit-learn package [21]. The data and source code that support the findings of this study are available from the corresponding author.

Results

Data Summary

A total of 51,474 individuals were screened between July 14, 2015 and April 21, 2018. Of these individuals, 2278 (4.6%) visited a physician within 2 months following the screening (see Figure S2 of Multimedia Appendix 1). Table S1 (Multimedia Appendix 1) displays the summary statistics for individuals who did and did not visit a doctor. Table 1 displays the summary statistics of all 2278 individuals in the final data set grouped by outcome. The average age was 50.6 years and 62% (1410/2278) of the patients were female. Both random blood glucose and blood pressure were notably high with averages of 167.8 mg/dL and 145/93 mmHg across all individuals, respectively.

Table 1. Summary of individual screening data (N=2278).[a]

Characteristic	Diabetes diagnosis		Hypertension diagnosis		All (N=2278)
	Positive (n=833)	Negative (n=1445)	Positive (n=1676)	Negative (n=602)	
Female, n (%)	496 (59.5)	914 (63.3)	1043 (62.2)	367 (61.0)	1410 (61.9)
Age (years), mean (SD)	52.9 (11.8)	49.3 (14.0)	52.3 (13.0)	45.8 (12.9)	50.6 (13.6)
Height (m), mean (SD)	1.6 (0.1)	1.6 (0.1)	1.6 (0.1)	1.6 (0.1)	1.6 (0.1)
Weight (kg), mean (SD)	66.5 (13.1)	62.1 (12.9)	64.5 (13.2)	61.5 (12.7)	63.7 (13.1)
BMI (kg/m²), mean (SD)	27 (5.4)	25.1 (5.4)	26.2 (5.5)	24.7 (5.2)	25.8 (5.5)
Waist circumference (cm), mean (SD)	93 (11.9)	90.4 (10.3)	91.9 (11.2)	90 (10.0)	91.4 (10.9)
Heart rate (per minute), mean (SD)	88 (12.2)	84 (11.6)	86 (12.1)	86.3 (11.2)	86.1 (11.9)
Random blood sugar (mg/dL), mean (SD)	233.6 (99.2)	129.9 (44.2)	166.7 (85.3)	171 (86.6)	167.8 (85.6)
Systolic blood pressure (mmHg), mean (SD)	144.4 (22.5)	144.6 (21.0)	147.6 (21.4)	136 (19.8)	144.6 (21.6)
Diastolic blood pressure (mmHg), mean (SD)	91.1 (12.2)	93.5 (12.0)	93.6 (12.0)	89.8 (11.8)	92.6 (12.1)
Urinations per night, mean (SD)	2 (1.2)	1.5 (0.9)	1.8 (1.1)	1.5 (1.0)	1.7 (1.1)
Parental diabetes, n (%)	244 (29.3)	297 (20.6)	410 (24.5)	131 (21.8)	541 (23.7)
Parental hypertension, n (%)	210 (25.2)	319 (22.1)	408 (24.3)	121 (20.1)	529 (23.2)
Dizziness, n (%)	124 (14.9)	106 (7.3)	184 (11.0)	46 (7.6)	230 (10.1)
Numbness, n (%)	153 (18.4)	117 (8.1)	224 (13.3)	46 (7.6)	270 (11.9)
Dry tongue, n (%)	162 (19.4)	97 (6.7)	216 (12.9)	43 (7.1)	259 (11.4)
Chest pain, n (%)	40 (4.8)	15 (1.0)	47 (2.8)	8 (1.3)	55 (2.4)
Current smoker, n (%)	67 (8.0)	100 (6.9)	129 (7.7)	38 (6.3)	167 (7.3)
Medication,[b] n (%)	578 (69.4)	413 (28.6)	863 (51.5)	128 (21.3)	991 (43.5)

[a] Diabetes and hypertension are doctor-reported diagnoses and correspond to our outcome (target) variable.

[b] Proportion of individuals currently using some type of medication, but no further details on type or reason for medication were collected.

Model Performance

Figure 1 displays the AUC distribution across all 250 test sets of the 10 models based on 5 machine learning approaches (5 for diabetes and 5 for hypertension). For diabetes, the random forest model had the highest average AUC value (mean [SD], 0.910 [0.001]), followed by logistic regression (mean [SD], 0.909 [0.001]), AdaBoost decision trees (mean [SD], 0.896 [0.002]), K-nearest neighbors (mean [SD], 0.857 [0.001]), and decision trees (mean [SD], 0.776 [0.005]). For hypertension, random forest (mean [SD], 0.792 [0.002]) performed slightly better than logistic regression (mean [SD], 0.776 [0.001]) and AdaBoost decision trees (mean [SD], 0.770 [0.003]). K-nearest neighbors (mean [SD], 0.705 [0.004]) and decision trees (mean [SD], 0.610 [0.01]) had poorer performance. All pairwise differences were found to be statistically significant. Given these results, we focused on the random forest model when comparing with baseline approaches for both diabetes and hypertension.

Figure 1. A comparison of area under the curve (AUC) distributions across 250 test sets between 5 risk stratification models for both diabetes and hypertension. The random forest model had the highest AUC for both diabetes (mean [SD], 0.910 [0.001]) and hypertension (mean [SD], 0.792 [0.002]). The upper and lower boundaries of the boxes correspond to the first and third quartiles, respectively. The line inside the box represents the median and the whiskers correspond to the minimum and the maximum of the distribution. KNN: K-nearest neighbors algorithm.

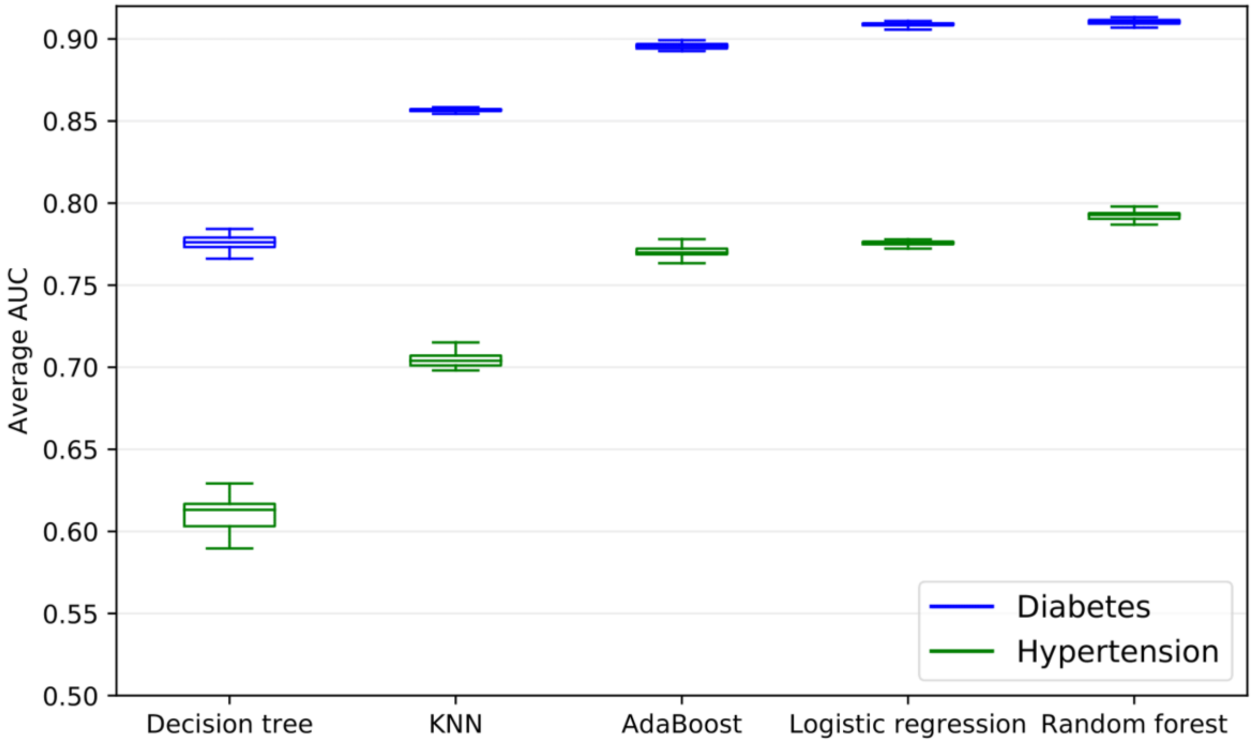
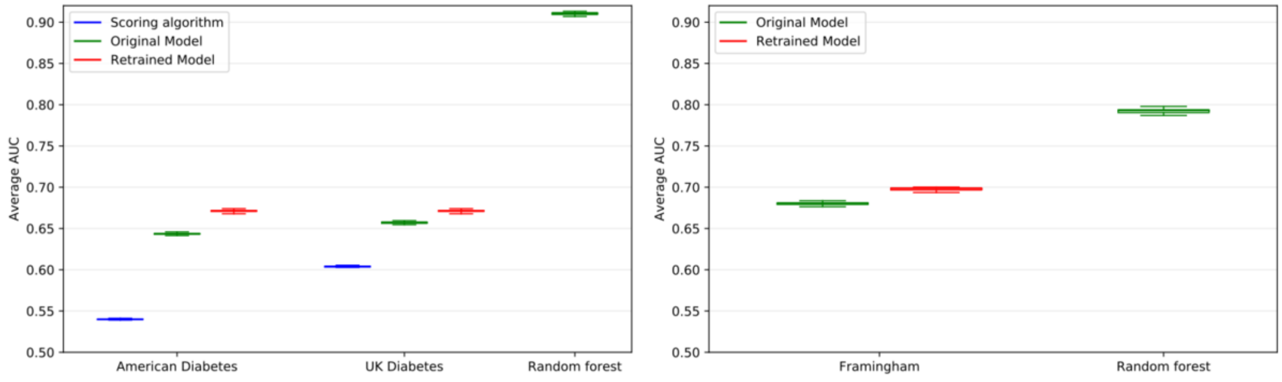


Figure 2 displays the AUC distribution for the random forest model and all baseline approaches for both diabetes and hypertension. For diabetes, the American Diabetes Association (ADA)–retrained (mean [SD], 0.671 [0.032]) and UK Diabetes–retrained (mean [SD], 0.671 [0.032]) performed best, followed by the UK–original model (mean [SD], 0.657 [0.031]) and the ADA–original model (mean [SD], 0.643 [0.033]). The ADA-scoring (mean [SD], 0.540 [0.021]) and UK-scoring (mean [SD], 0.604 [0.031]) methods performed considerably worse.

For hypertension, Framingham-retrained (mean [SD], 0.698 [0.037]) performed slightly better than the Framingham–original model (mean [SD], 0.680 [0.036]). The random forest model significantly outperformed all baseline approaches. Retraining the baseline models using our data provided a statistically significant increase in their accuracy. Nevertheless, the accuracy of these retrained models was still lower than the accuracy of our random forest model.

Figure 2. A comparison of the area under curve (AUC) distributions for the random forest model and all baseline approaches. A. For diabetes, the AUC of the random forest model improved upon the best baseline approach (UK diabetes–retrained) by 0.239 (0.910 vs 0.671, *P*<.001). B. For hypertension, the AUC of the random forest model improved upon the best baseline approach (Framingham-retrained) by 0.095 (0.792 vs 0.697, *P*<.001).



A. Diabetes

B. Hypertension

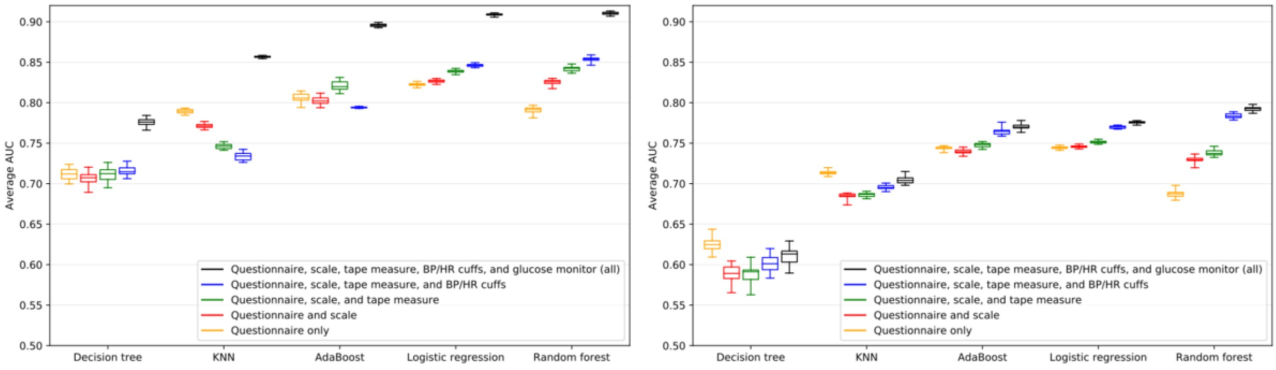
To visualize the trade-off between false positives and false negatives, Figure S3 (Multimedia Appendix 1) displays receiver operating characteristic curves from a single randomly selected test set (out of 250) of the random forest model and baseline approaches for both diabetes and hypertension. For a fixed screening specificity of 0.9, the random forest model was able to reduce the false negative rate, on average from 0.79 to 0.17 for diabetes and from 0.72 to 0.50 for hypertension. In other words, the random forest model can reduce the number of false negatives by 620 patients per 1000 screenings for diabetes and 220 patients per 1000 screenings for hypertension.

Figure S4 (Multimedia Appendix 1) displays the normalized feature importance extracted from the random forest model. As expected, blood sugar was the most important feature for diabetes risk prediction, but several other features including many self-reported symptoms (eg, urination, dry tongue) provided predictive power. For hypertension, systolic blood pressure was the most important, followed closely by blood sugar, which is not used for hypertensive risk prediction in high-income countries, even though there is a known link between these diseases.

Model Performance as a Function of Data Availability

Figure 3 displays the AUC distributions for the 5 risk stratification models and the 5 different features sets. For both diabetes and hypertension, a random forest model with access to only questionnaire-type features was able to capture 87% of the AUC obtained from a model with access to all features. We found that the use of a glucose monitor had the largest impact on diabetes model performance, increasing the average AUC by more than 0.05 for all models, while the use of a blood pressure/heart rate cuff had the largest impact on hypertensive risk prediction, increasing the average AUC by a mean of 0.05 across all models. See Figure S5 (Multimedia Appendix 1) for a visualization of the trade-off between false positives and false negatives.

Figure 3. Area under the curve (AUC) distributions for 5 risk stratification models and 5 different features sets. A. Including a glucose monitor had the largest effect on diabetes risk stratification, increasing average AUC by more than 0.05 for all models. B. Including blood pressure/heart rate cuffs had the largest effect on hypertension risk stratification, increasing the average AUC by up to 0.05. BP: blood pressure; HR: heart rate; KNN: K-nearest neighbors algorithm.



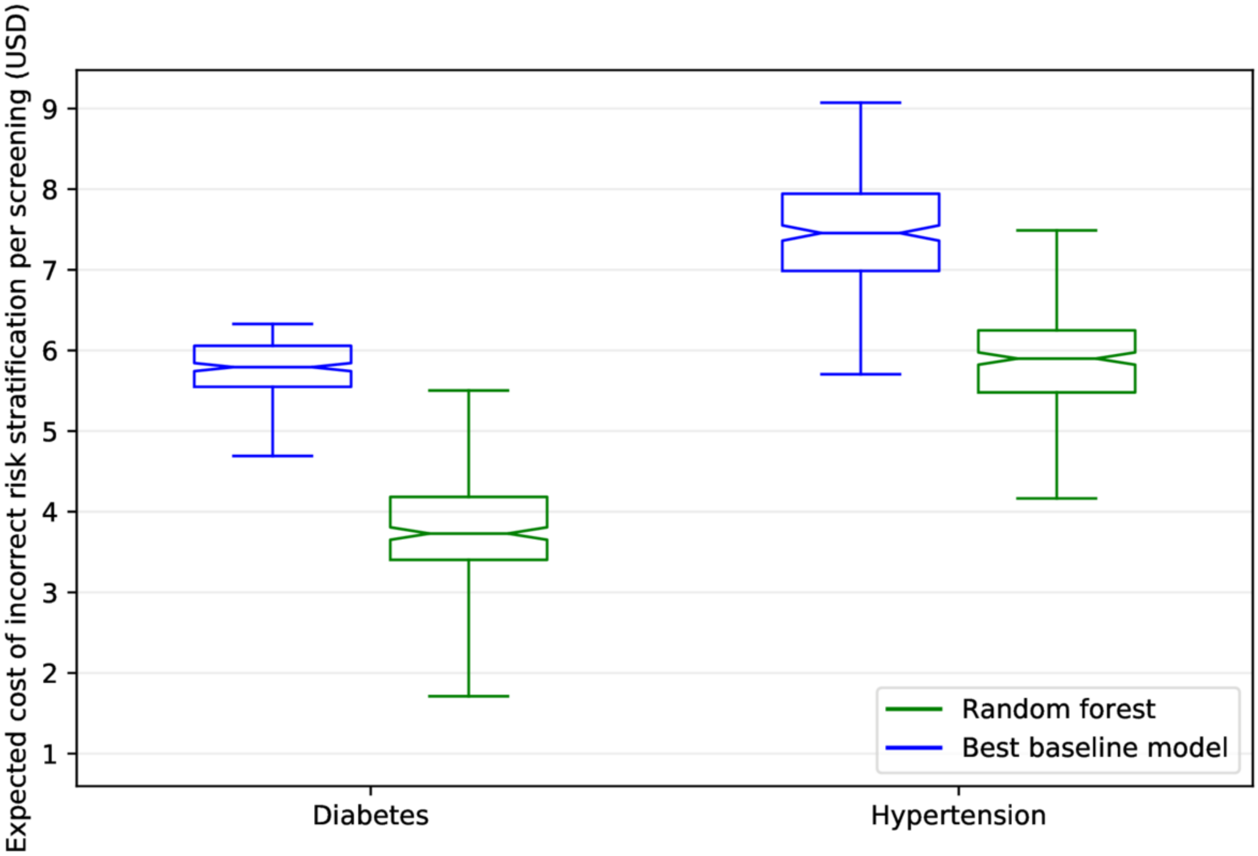
A. Diabetes

B. Hypertension

Cost Analysis

Figure 4 displays the expected cost of incorrect risk stratification per screening for the random forest model and the best baseline approach for diabetes (UK Diabetes–retrained) and hypertension (Framingham-retrained). For the baseline models, the expected cost of incorrect risk stratification per screening was US $5.76 and US $7.47 for diabetes and hypertension, respectively. The random forest model was able to reduce the expected cost of incorrect risk stratification per screening by US $1.99 (or 35%) for diabetes and US $1.60 (or 21%) for hypertension. All cost reductions were found to be statistically significant. Figure S6 (Multimedia Appendix 1) displays the results of our sensitivity analysis.

Figure 4. A comparison of the expected cost of incorrect risk stratification per screening across 250 test sets between the random forest and best baseline approach for both diabetes and hypertension. The random forest reduced the expected cost of incorrect risk stratification by US $1.99 per screening for diabetes (US $5.76 vs US $3.77, *P*<.001) and by US $1.60 per screening for hypertension (US $7.47 vs US $5.87, *P*<.001). The upper and lower boundaries of the boxes correspond to the first and third quartiles, respectively. The line inside the box represents the median, the whiskers correspond to the minimum and the maximum of the distribution, and the notches in the box represent the 95% confidence interval around the median.



Discussion

This study developed risk stratification models to predict the short-term (2-month) risk in a resource-limited setting for both diabetes and hypertension. Our primary analysis demonstrated that models from high-income countries do not generalize well to the low- and middle-income countries. In particular, the random forest model had the highest prediction accuracy for both diseases and was able to outperform the best baseline approach in terms of AUC by 35.5% for diabetes and 13.5% for hypertension. Our secondary analysis found that risk stratification can be accurately performed with limited data. A random forest model with access to only questionnaire-type features was able to capture 87% of the AUC obtained from a model with access to all features, suggesting that diabetes and hypertension risk stratification can be accurately conducted in extremely resource-limited settings. Although there are circumstances where advanced measurements may be required, eliminating the need for the corresponding tools means that community health workers require less training and can travel with fewer devices.

The observed performance difference between the baseline approaches and our models can be attributed to 3 improvements. First, our models were designed for short-term risk prediction, while the baseline models were designed for long-term prediction. Even though we retrained the baseline models with our data, the features included in the models were selected for long-term prediction. For example, none of the baseline models included self-reported symptoms (eg, dry tongue, urination), which may be more suitable for short-term prediction. Second, our models include additional features not used by the baseline approaches that may provide additional insight into the social, lifestyle, and genetic differences in the population. For example, none of the risk scores from high-income countries use self-reported symptoms or random blood glucose. Although random blood glucose is not typically used in high-income settings where HbA_{1c} is preferred, it is often captured by community-based screening programs due to its operational simplicity (eg, no fasting required). For diabetes, random blood glucose was the most important feature and increased the AUC by 0.13, while for hypertension, random blood glucose was the second most important feature (see Figure S4 of Multimedia Appendix 1) and also led to an AUC increase. Third, we believe that the advanced machine learning models allowed us to extract maximum value from the small sample size and simple features available to us, whereas simple models with advanced features and large data sets may be equally effective in high-income settings.

As a by-product of our analysis, we externally validated the previously developed baseline approaches by using India-specific data. Although many of these models have been externally validated in a variety of settings, they have not been compared using India-specific data [10,11]. For example, the Framingham model for hypertensive risk has been validated in 7 countries with an average AUC of 0.80 (range 0.73-0.84) [11]. Our results show that the model is not as effective in India, where it had an average AUC of 0.70 after being retrained using local data. It is challenging to determine why the model performed poorly, but we believe that it may be due to subtle differences in the at-risk population, which manifest in the features selected by the model. Overall, our validation and comparison of baseline models highlights the importance of developing risk prediction models specifically for the low- and middle-income countries of interest.

The translation of our findings to the design and implementation of nation-wide screening programs must carefully consider costs, field accessibility, and disease management. The results of our secondary analysis indicate that the most impactful features (blood glucose, blood pressure, and heart rate) are measured using the most expensive field equipment (glucose monitor and blood pressure/heart rate cuffs). Even though these devices are more expensive, we find that including glucose monitors for diabetes screening and heart rate/blood pressure cuffs for hypertension screening can reduce the expected cost of incorrect risk stratification by US $1.35 and US $0.70, respectively (see Figure S7 of Multimedia Appendix 1). A formal cost-effectiveness analysis is needed to determine whether the gain in accuracy (and subsequent reduction is risk stratification cost) is worth the capital investment required to purchase glucose monitors and heart rate/blood pressure cuffs in low-resource settings.

There is also an important cost-tradeoff between a high false positive rate and a high false negative rate, which is determined by the discriminant threshold used to stratify patients into risk categories. Research suggests that the financial cost of a false positive is minimal (US $7 for diabetes and US $15 for hypertension) compared to that of a false negative (US $288 for diabetes and US $45 for hypertension) [22]. Our results demonstrated that the random forest model can reduce the number of false negatives by 620 patients per 1000 screenings for diabetes and 220 patients per 1000 screenings for hypertension. Extrapolating these results to a nationwide screening program in India that screens 600 million people [23] could save approximately US $1.19 billion for diabetes and US $960 million for hypertension by reducing the false negatives. In the next decade, the central government of India is planning to spend significant resources on noncommunicable disease screening programs [8] and our models can be leveraged by these screening programs to effectively utilize limited resources by improving risk stratification accuracy.

Despite the complex nature of our models, they can be easily implemented and computed into handheld tablets (or other mobile health devices) carried by community health workers without the need for a simplified, hand-computable risk score, which means we can provide the most accurate prediction without any extra effort or calculations by the community health workers. Furthermore, mobile health applications have demonstrated the ability to increase access to health care for low-income populations and to improve the capacity of the existing health systems [24]. Future research is needed to understand how to best integrate and present the risk stratification results into the community health worker workflow.

It is important to note that screening is only the first step to reducing the burden of noncommunicable diseases. Once high-risk patients are identified, they need to be linked to appropriate care and put on a disease management plan [25]. Linking patients to care and initiating disease management is a nontrivial process and governments need to carefully design nationwide disease management plans because otherwise, screening programs are unlikely to have the desired impact. Therefore, an important direction for future research includes studying the effect of screening programs on population health outcomes in the presence of current and enhanced linkages to care and disease management plans.

Our work has several limitations. First, we did not have access to an external validation set from a different study population (eg, rural slums, different state or country) to test our models. Second, our data displays a clear selection bias toward sicker patients visiting a physician within 2 months (see Table S1 of Multimedia Appendix 1). From a risk stratification perspective, the selection bias toward sicker individuals makes the problem more difficult because the model must discriminate between similar individuals. In other words, we need to identify those who actually have diabetes or hypertension from a pool of individuals who all appear to be at high risk. Finally, the differences in disease prevalence and overall health between our sample and the National Family Health Survey, Hyderabad suggest that, if applied broadly, our model may experience data shifting, which occurs when the training data differs from the application data [26]. See Table S3 of Multimedia Appendix 1 for a comparison of our data sample with the urban sample from India's National Family Health Survey. Data shifting can negatively impact accuracy (similar to how the models from the United States and the United Kingdom performed poorly in India) and future research is needed to test our models in other settings.

In conclusion, this study found that a machine learning–based risk stratification model tailored to data collected by community-based screening programs can significantly improve risk stratification accuracy for both diabetes and hypertension in low-resource settings. Researchers and international organizations have proposed machine learning as a game changer in global health, [27-29] but there is limited documented evidence that machine learning can be effectively utilized in the resource-limited settings indicative of global health projects [30]. This study adds evidence to support machine learning in global health by quantitatively demonstrating the benefit of using these models in a novel resource-limited context.

Acknowledgments

We thank NanoHealth for providing data and feedback to our approach. JJB was supported by a Natural Sciences and Engineering Research Council of Canada Michael Smith Foreign Study Supplement, the Max Institute of Healthcare Management, and the Indian School of Business. The data and source code that support the findings of this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[PDF File (Adobe PDF File), 579 KB - jmir_v23i1e20123_app1.pdf]

References

1. Noncommunicable diseases: key facts. World Health Organization. 2018. URL: http://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2020-05-11]
2. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
3. Anchala R, Kannuri NK, Pant H, Khan H, Franco OH, Di Angelantonio E, et al. Hypertension in India. Journal of Hypertension 2014;32(6):1170-1177. [doi: 10.1097/hjh.0000000000000146]
4. Bansode B, Nagarajan R. Diabetes: A review of awareness, comorbidities, and quality of life in India. Journal of Social Health and Diabetes 2018 Nov 23;05(02):77-82. [doi: 10.1055/s-0038-1676248]
5. Herman WH, Ye W, Griffin SJ, Simmons RK, Davies MJ, Khunti K, et al. Early Detection and Treatment of Type 2 Diabetes Reduce Cardiovascular Morbidity and Mortality: A Simulation of the Results of the Anglo-Danish-Dutch Study of Intensive Treatment in People With Screen-Detected Diabetes in Primary Care (ADDITION-Europe). Diabetes Care 2015 Aug;38(8):1449-1455 [FREE Full text] [doi: 10.2337/dc14-2459] [Medline: 25986661]
6. Fukuma S, Ikenoue T, Saito Y, Yamada Y, Saigusa Y, Misumi T, et al. Lack of a bridge between screening and medical management for hypertension: health screening cohort in Japan. BMC Public Health 2020 Sep 17;20(1):1419 [FREE Full text] [doi: 10.1186/s12889-020-09532-5] [Medline: 32943038]
7. Zeltner T, Riahi F, Huber J. Acute and Chronic Health Challenges in Sub-Saharan Africa: An Unfinished Agenda. In: Groth H, May JF, editors. Africa's Population: In Search of a Demographic Dividend. Cham, Switzerland: Springer International Publishing; 2017:283-297.
8. National Health Policy. Ministry of Health and Welfare, Government of India. 2017. URL: https://www.nhp.gov.in/nhpfiles/national_health_policy_2017.pdf [accessed 2020-05-11]
9. Jeet G, Thakur JS, Prinja S, Singh M. Community health workers for non-communicable diseases prevention and control in developing countries: Evidence and implications. PLoS One 2017;12(7):e0180640 [FREE Full text] [doi: 10.1371/journal.pone.0180640] [Medline: 28704405]
10. Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AMW, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. BMJ 2012 Sep 18;345:e5900 [FREE Full text] [doi: 10.1136/bmj.e5900] [Medline: 22990994]
11. Sun D, Liu J, Xiao L, Liu Y, Wang Z, Li C, et al. Recent development of risk-prediction models for incident hypertension: An updated systematic review. PLoS One 2017;12(10):e0187240 [FREE Full text] [doi: 10.1371/journal.pone.0187240] [Medline: 29084293]
12. Gujral UP, Pradeepa R, Weber MB, Narayan KV, Mohan V. Type 2 diabetes in South Asians: similarities and differences with white Caucasian and other populations. Ann N Y Acad Sci 2013 Apr;1281:51-63 [FREE Full text] [doi: 10.1111/j.1749-6632.2012.06838.x] [Medline: 23317344]
13. Cappuccio F. Ethnicity and cardiovascular risk: variations in people of African ancestry and South Asian origin. J Hum Hypertens 1997 Sep;11(9):571-576. [doi: 10.1038/sj.jhh.1000516] [Medline: 9364274]
14. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007 May 28;167(10):1068-1074. [doi: 10.1001/archinte.167.10.1068] [Medline: 17533210]
15. Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AMW, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. BMJ 2012 Sep 18;345:e5900 [FREE Full text] [doi: 10.1136/bmj.e5900] [Medline: 22990994]

16. Clark CM, Snyder JW, Meek RL, Stutz LM, Parkin CG. A systematic approach to risk stratification and intervention within a managed care environment improves diabetes outcomes and patient satisfaction. Diabetes Care 2001 Jun;24(6):1079-1086. [doi: 10.2337/diacare.24.6.1079] [Medline: 11375374]
17. Census of India. Government of India. 2011. URL: http://censusindia.gov.in/ [accessed 2020-05-11]
18. World Bank Team Visits Hyderabad Slums. The Times of India. 2012. URL: https://timesofindia.indiatimes.com/city/hyderabad/World-bank-team-visits-Hyderabad-slums/articleshow/14036371.cms [accessed 2020-05-11]
19. Deo S, Singh P. Evaluation of a novel community-based chronic disease management solution. Max Institute of Healthcare Management. 2018. URL: https://www.isb.edu/en/research-thought-leadership/research-centres-institutes/max-institute-of-health-care-management/reports.html [accessed 2020-05-11]
20. Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1945 Dec;1(6):80. [doi: 10.2307/3001968]
21. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011. URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf [accessed 2020-05-11]
22. Chatterjee R, Narayan KMV, Lipscomb J, Jackson SL, Long Q, Zhu M, et al. Screening for diabetes and prediabetes should be cost-saving in patients at high risk. Diabetes Care 2013 Jul;36(7):1981-1987 [FREE Full text] [doi: 10.2337/dc12-1752] [Medline: 23393215]
23. Basu S, Millett C, Vijan S, Hayward RA, Kinra S, Ahuja R, et al. The health system and population health implications of large-scale diabetes screening in India: a microsimulation model of alternative approaches. PLoS Med 2015 May;12(5):e1001827; discussion e1001827 [FREE Full text] [doi: 10.1371/journal.pmed.1001827] [Medline: 25992895]
24. Sondaal SFV, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the Effect of mHealth Interventions in Improving Maternal and Neonatal Care in Low- and Middle-Income Countries: A Systematic Review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
25. Venkataraman K, Kannan AT, Mohan V. Challenges in diabetes management with particular reference to India. Int J Diabetes Dev Ctries 2009 Jul;29(3):103-109 [FREE Full text] [doi: 10.4103/0973-3930.54286] [Medline: 20165646]
26. Sugiyama M, Lawrence N, Schwaighofer A. Dataset Shift in Machine Learning. Cambridge, MA: The MIT Press; 2008.
27. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
28. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019 Mar;25(3):433-438. [doi: 10.1038/s41591-018-0335-9] [Medline: 30742121]
29. Ravizza S, Huschto T, Adamov A, Böhm L, Büsser A, Flöther FF, et al. Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. Nat Med 2019 Jan;25(1):57-59. [doi: 10.1038/s41591-018-0239-8] [Medline: 30617317]
30. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018;3(4):e000798 [FREE Full text] [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]

Abbreviations

ADA: American Diabetes Association
AUC: area under the receiver operating characteristic curve
HbA_{1c}: glycated hemoglobin

Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by C Zhang, S Baxter, L Pan; comments to author 23.10.20; revised version received 17.12.20; accepted 19.12.20; published 21.01.21.

Please cite as:
Boutilier JJ, Chan TCY, Ranjan M, Deo S
Risk Stratification for Early Detection of Diabetes and Hypertension in Resource-Limited Settings: Machine Learning Analysis
J Med Internet Res 2021;23(1):e20123
URL: http://www.jmir.org/2021/1/e20123/
doi:10.2196/20123
PMID:33475518

Original Paper

Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study

Xiangmin Fan[1*], PhD; Daren Chao[2*], BSc; Zhan Zhang[3*], PhD; Dakuo Wang[4*], PhD; Xiaohua Li[5], MSc; Feng Tian[1*], PhD

[1]The Institute of Software, Chinese Academy of Sciences, Beijing, China
[2]Department of Computer Science, University of Toronto, Toronto, ON, Canada
[3]School of Computer Science and Information Systems, Pace University, New York, NY, United States
[4]IBM Research, Cambridge, MA, United States
[5]Zuoshouyisheng Inc, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Zhan Zhang, PhD
School of Computer Science and Information Systems
Pace University
1 Pace Plaza
New York, NY, 10078
United States
Phone: 1 9147733254
Email: zzhang@pace.edu

Abstract

Background: Artificial intelligence (AI)-driven chatbots are increasingly being used in health care, but most chatbots are designed for a specific population and evaluated in controlled settings. There is little research documenting how health consumers (eg, patients and caregivers) use chatbots for self-diagnosis purposes in real-world scenarios.

Objective: The aim of this research was to understand how health chatbots are used in a real-world context, what issues and barriers exist in their usage, and how the user experience of this novel technology can be improved.

Methods: We employed a data-driven approach to analyze the system log of a widely deployed self-diagnosis chatbot in China. Our data set consisted of 47,684 consultation sessions initiated by 16,519 users over 6 months. The log data included a variety of information, including users' nonidentifiable demographic information, consultation details, diagnostic reports, and user feedback. We conducted both statistical analysis and content analysis on this heterogeneous data set.

Results: The chatbot users spanned all age groups, including middle-aged and older adults. Users consulted the chatbot on a wide range of medical conditions, including those that often entail considerable privacy and social stigma issues. Furthermore, we distilled 2 prominent issues in the use of the chatbot: (1) a considerable number of users dropped out in the middle of their consultation sessions, and (2) some users pretended to have health concerns and used the chatbot for nontherapeutic purposes. Finally, we identified a set of user concerns regarding the use of the chatbot, including insufficient actionable information and perceived inaccurate diagnostic suggestions.

Conclusions: Although health chatbots are considered to be convenient tools for enhancing patient-centered care, there are issues and barriers impeding the optimal use of this novel technology. Designers and developers should employ user-centered approaches to address the issues and user concerns to achieve the best uptake and utilization. We conclude the paper by discussing several design implications, including making the chatbots more informative, easy-to-use, and trustworthy, as well as improving the onboarding experience to enhance user engagement.

(J Med Internet Res 2021;23(1):e19928) doi:10.2196/19928

KEYWORDS

self-diagnosis; chatbot; conversational agent; human–artificial intelligence interaction; artificial intelligence; diagnosis; case study; eHealth; real world; user experience

Introduction

Background

Medical consultation has historically been conducted between patients and their physicians during clinical encounters. However, some barriers may hinder effective communication in a clinical setting. For example, patients and their caregivers often face great challenges obtaining timely medical advice and information from health care providers due to the long wait time for an appointment [1,2]. The increasing demand for health care services also places a large burden on health care providers due to the shortage of medical professionals [3,4]. Medical professionals, therefore, have to overcome a range of temporal, geographical, and organizational barriers to provide a high quality of health care to patients [5]. Even more concerning is that health care infrastructures (the underlying foundation that supports the operations of a health care system) are complex and fragmented in many countries [6]. The scarceness and imbalanced distribution of health care resources (eg, facilities and medical professionals) often impede people in rural areas from accessing health care services and obtaining professional medical advice in a timely and effective manner [7,8]. Therefore, patients often use other mechanisms, such as online health forums and “Ask the Doctor” services, to seek medical help by asking questions and receiving answers from peers and health professionals who are willing to share their knowledge and opinions [1,9-12]. Even so, patients may still not be able to get instant responses using these online platforms [13,14]. Furthermore, there is much inaccurate information online, which may easily mislead patients [15].

With the advances in artificial intelligence (AI) technology in recent years, there is an opportunity to tackle the challenges and barriers faced by patients in seeking timely health information and to reduce the burdens posed on medical professionals [16,17]. That is, AI-driven intelligent systems, such as health chatbots, have emerged to support patients seeking medical advice irrespective of time and place [18]. These chatbots can provide live feedback to help patients get an overview of their symptoms, become aware of their illness, triage and manage their conditions, and ultimately improve their health [19-22]. Such chatbots act as a virtual conversational agent mimicking human interactions and offering medical advice (eg, diagnostic suggestions) directly to patients in a timely and cost-effective manner. In this way, health chatbots provide a form of triage into the health care system and become the first point of contact for health. While this technology is still in its developmental phase, chatbot systems could potentially alter the landscape of health care by increasing access to health care services, enhancing patient-centered care, and reducing unnecessary clinical visits [23,24].

Despite these potential benefits, similar to many other mobile health (mHealth) applications, chatbot systems have been inadequately adopted by those who might benefit most from this novel technology [4,25]. It is therefore important to examine how to design health chatbots to increase user adoption and engagement. Furthermore, prior work has focused primarily on developing advanced algorithms to improve the accuracy and effectiveness of chatbots’ diagnoses [21,26,27]. Few studies, however, have focused on the use of health chatbots in the real world [25]. More specifically, little is known about the “who, how often, and why” of chatbot use, what barriers and issues exist in using this novel technology, and how to overcome the barriers. This research gap is caused partly by the lack of large-scale deployment of health chatbots [28]. That is, most studies only examined the use of health chatbots in controlled settings rather than in real-world settings, where users interact with chatbots on their own.

To bridge this knowledge gap, we examined one widely deployed chatbot in China—DoctorBot [29]—which had attracted hundreds of thousands of users by the time this study took place. The large-scale deployment of DoctorBot provides us with a unique opportunity to gain an in-depth, empirical understanding of how people use health chatbots in the real world and what barriers hinder the delivery of these novel services. To the best of our knowledge, this is the first study that examined these issues using large-scale, real usage data. More specifically, we took a data-driven approach to analyze the system log of DoctorBot, which consisted of 47,684 consultation sessions initiated by 16,519 users between September 2018 and March 2019. Through this study, we make three contributions. First, we present a detailed analysis of how people use an AI-driven self-diagnosis chatbot, which remains understudied in the health informatics community. Second, we report issues and barriers that hinder the effective use of health chatbots. Third, our results can shed light on how to better design health chatbots to optimize user experience and achieve the best uptake and utilization.

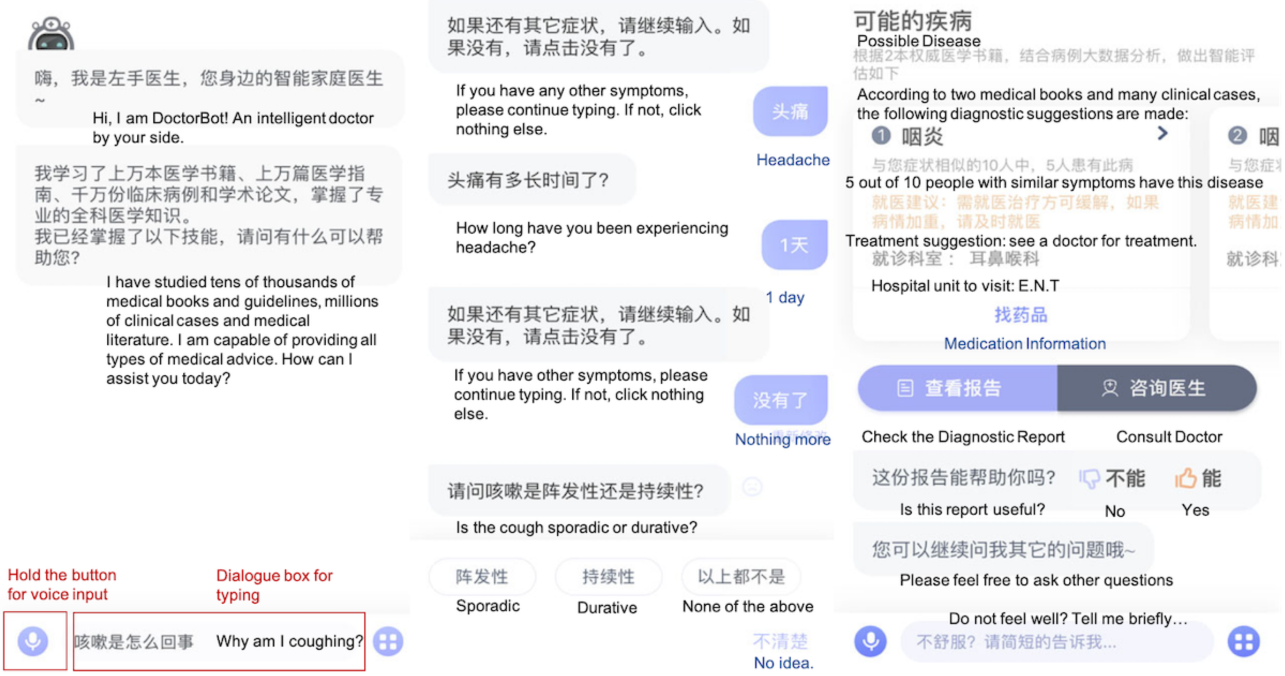
Features of DoctorBot

DoctorBot is an AI-driven, mobile-based medical consultation platform [29]. It utilizes large data sets (eg, numerous medical literature and clinical cases) and state-of-the-art AI techniques (eg, deep learning and knowledge graphs) to process users’ inquiries and provide personalized medical advice. Users can interact with DoctorBot by typing information into a chatbox and/or recording a voice message to express their health concerns (the voice message can be converted into text in real time). DoctorBot provides different health services to users, such as self-diagnosis, drug use instructions, diet suggestions, and so forth. Among those, self-diagnosis is one of the most popular and demanding services. Users can explain their health concerns to the chatbot and receive medical advice (eg, diagnostic suggestions and treatment options) to make informed decisions. Given the current heated debate on the readiness and usefulness of self-diagnosis chatbots [30,31], we chose to focus on the use of the self-diagnosis feature in this study.

A typical consultation session starts with a prompt for the user to describe their main symptom or chief complaint (Figure 1, left). After being prompted, the user provides input, such as, “Why am I coughing?” This inquiry triggers DoctorBot to take the initiative and ask the user a series of questions related to the symptoms (eg, “How long has the cough been present?”). The chatbot may also inquire if the user is experiencing other relevant symptoms (Figure 1, middle). For example, if a user indicates that he/she has been coughing, the chatbot would consecutively ask if the user is also experiencing a sore throat, shortness of breath, and so forth. If the answer is “yes,” the chatbot then asks the user more questions about each relevant symptom. Finally, DoctorBot asks the user to provide his/her medical history to conclude the consultation.

When a consultation is complete, DoctorBot generates a report detailing potential diagnoses, prediction confidence, treatment options, and which hospital unit (eg, cardiology, urology) to visit (Figure 1, right). Such information is expected to help users decide when, where, and whether or not to seek further medical help. It is also worth noting that DoctorBot explicitly instructs users to use the diagnosis for reference only, in light of AI liability issues and medical ethics [32].

Figure 1. DoctorBot interfaces (all Chinese dialogue was translated into English in this figure). Left: the starting point of a consultation; after greetings, the user is prompted to explain health concerns. Middle: DoctorBot asks the user a series of symptom-related questions; users can select from the offered options at the bottom to provide an answer. Right: a report detailing potential diagnoses (ranked by their possibility) and other medical advice is provided to the user. E.N.T: ear, nose, and throat clinic.



嗨，我是左手医生，您身边的智能家庭医生~

Hi, I am DoctorBot! An intelligent doctor by your side.

我学习了上万本医学书籍、上万篇医学指南、千万份临床病例和学术论文，掌握了专业的全科医学知识。
我已经掌握了以下技能，请问有什么可以帮助您？

I have studied tens of thousands of medical books and guidelines, millions of clinical cases and medical literature. I am capable of providing all types of medical advice. How can I assist you today?

咳嗽是怎么回事 Why am I coughing?

如果还有其它症状，请继续输入。如果没有，请点击没有了。

If you have any other symptoms, please continue typing. If not, click nothing else.

头痛 Headache

头痛有多长时间了？

How long have you been experiencing headache?

1天 1 day

如果还有其它症状，请继续输入。如果没有，请点击没有了。

If you have other symptoms, please continue typing. If not, click nothing else.

没有了 Nothing more

请问咳嗽是阵发性还是持续性？

Is the cough sporadic or durative?

可能的疾病
Possible Disease

根据2本权威医学书籍，结合病例大数据分析，做出智能评估如下

According to two medical books and many clinical cases, the following diagnostic suggestions are made:

① 咽炎

与您症状相似的10人中，5人患有此病

5 out of 10 people with similar symptoms have this disease

就医建议：需就医治疗方可缓解，如果病情加重，请及时就医

Treatment suggestion: see a doctor for treatment.

就诊科室： 耳鼻喉科

Hospital unit to visit: E.N.T

这份报告能帮助你吗？

Is this report useful?

您可以继续问我其它的问题哦~

Please feel free to ask other questions

不舒服？请简短的告诉我...

Do not feel well? Tell me briefly…

Methods

Data Collection and Processing

To understand the use of self-diagnosis chatbots in the real world, we took a data-driven approach to analyze the system log of DoctorBot collected between September 2018 and March 2019. The data set consisted of 47,684 consultation sessions initiated by 16,519 users over the 6-month period. The log of each consultation session included a session ID, the user’s general information (eg, age, gender), consultation details (eg, time stamps, and conversation content between DoctorBot and the user), a diagnostic report automatically generated by the chatbot after the consultation was completed, and user feedback (if the user voluntarily provided feedback regarding the usefulness of the chatbot at the end).

The first author worked closely with the chatbot company to determine the type, amount, and format of usage data that needed to be extracted to fulfill the study’s purpose. To protect users’ privacy, all identifiable information (eg, phone number) was removed from the data set by the chatbot company before sending it to the researchers. Therefore, the data we analyzed was completely anonymized. Moreover, the users consented at the point of registration that researchers were allowed to analyze their usage data for research purposes. The study procedure was approved by the first author’s university’s Institutional Review Board.

The initial data set had some noisy data; for example, one consultation session could be stored as two separate sessions, and two consultation sessions could be merged. Therefore, we preprocessed the data by splitting the sticky conversations and spliced the broken conversations. After data processing, the research team examined the entire data set to ensure the accuracy and appropriateness of the data format. Since the study context was based in China, the content of the dialogues between the chatbot and users was in Chinese. To ensure the validity of the data, we decided to analyze the content in its original language. We used the Jieba [33] word segmentation library to segment the user input to extract semantic information.

Data Analysis

To understand who uses chatbots, how often, and why, we first performed descriptive statistical analysis on the entire data set, focusing on user characteristics (eg, gender and age) and general patterns of chatbot use (eg, duration, frequency, and purpose of use). This analysis helped us not only to obtain a general overview of chatbot use but also to identify an interesting scenario—only 30,710 of 47,684 consultation sessions were completed, with the remaining 16,974 sessions being terminated by users before the chatbot reached a diagnosis. Therefore, we decided to split the entire data set into "completed" and "uncompleted" consultation sessions for further analysis. In particular, we used content analysis [34] and statistical analysis in combination to analyze both types of sessions to investigate the issues and barriers that may occur during the interactions between DoctorBot and users. More specifically, because an uncompleted consultation session may suggest that a user encountered barriers in using the chatbot, we investigated the exact moment when a consultation was terminated to understand what questions were asked by DoctorBot and how much time users spent on answering these questions. We also performed statistical analysis (ie, logistic regression and principal component analysis [PCA]) to explore the characteristics of uncompleted consultations. Examining these aspects could help us gain a preliminary understanding of the factors that could potentially lead to user dropout. In addition to the analysis of uncompleted sessions, two researchers performed a content analysis on 3000 completed sessions that were randomly selected to examine interaction issues that may exist in the use of the chatbot, such as miscommunication between DoctorBot and users. The content analysis was augmented by statistical analysis to further explore the influencing factors on the emerged interaction issues. Finally, as DoctorBot usually prompts the user to provide feedback (eg, positive versus negative experience) toward the end of the consultation, we performed a content analysis on the user feedback to identify user concerns and issues related to the use of DoctorBot and explored what factors could contribute or lead to these user concerns.

Results

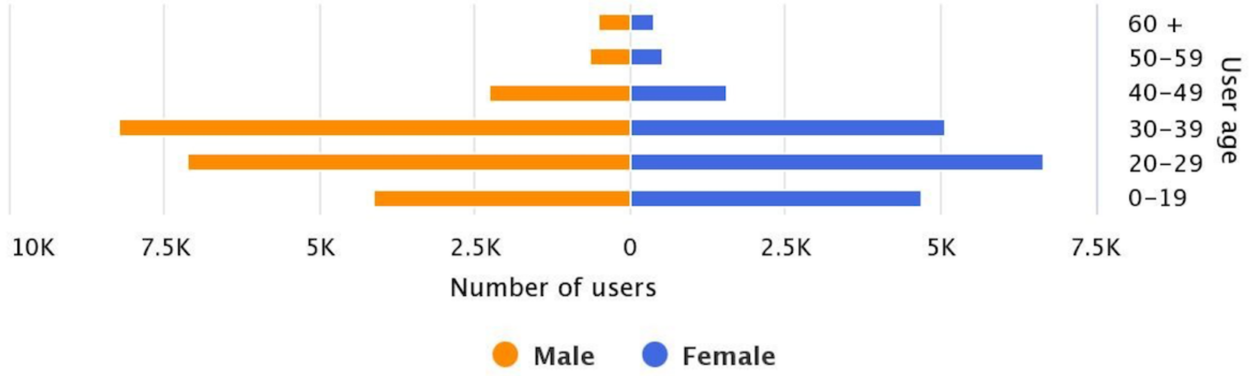
General Patterns of Chatbot Use

We first report the results that emerged from the statistical analysis on the entire data set. This analysis helped us understand several patterns of the large-scale use of DoctorBot, including who used it, the length of each consultation, how often users used the application, and what health concerns users had queried about.

Who Used DoctorBot?

During the data collection period (September 2018 to March 2019), 16,519 users interacted with DoctorBot to present health concerns. Our analysis of users' demographic information allowed us to obtain an understanding of the user groups and their characteristics. As Figure 2 shows, more than one-half of the consultations (9052/16,519, 54.80%) were initiated by male users, and the mean ages of male and female users were 30 years and 27 years, respectively. In particular, the majority of DoctorBot users were between ages 20 and 39 years, which may be due to the relatively high technology proficiency and willingness to use advanced technologies (ie, intelligent chatbot) among this age group. Many users in the 20 to 39 years age group also consulted the application concerning their childrens' illnesses (eg, "May I ask the reason why a 6-year-old girl often vomits?"). In contrast, only a small number of older adults (aged >60 years) or their caregivers (n=885) attempted to use the chatbot. This finding aligns with previous research showing that older adults are lagging in the adoption of novel health-related technologies [35]. To bridge the gap, more research is needed to examine how to design advanced health technologies (eg, chatbots) tailored to the aging population.

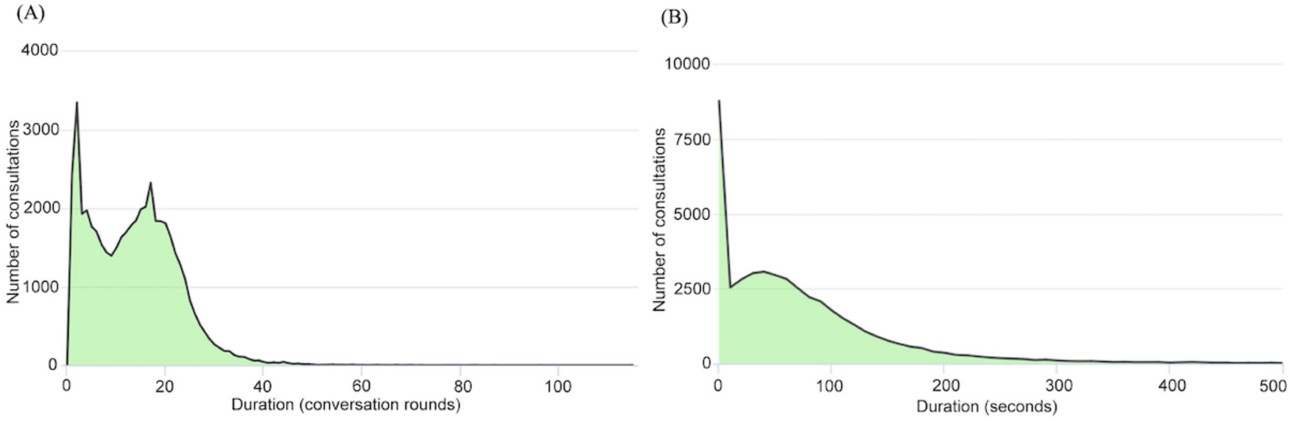
Figure 2. Characteristics of DoctorBot users. K: thousand.



What Were the Length and Frequency of Chatbot Sessions?

As shown in Figure 3, the duration of each consultation varied. Many consultation sessions only lasted a few conversation rounds or seconds, whereas some others took more than 20 conversation rounds or 5 minutes. This finding highlights that many users only interacted with the chatbot for a brief time, which may not have allowed them to complete a consultation session.

Figure 3. (A) Consultation duration in conversation rounds. (B) Consultation duration in seconds.

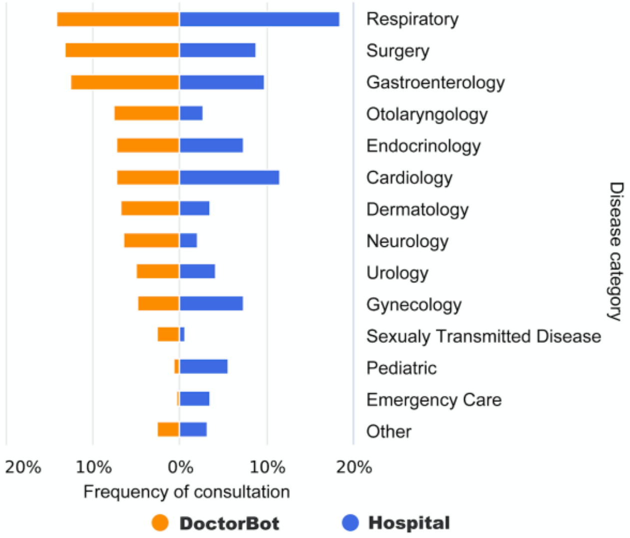


Regarding the frequency of chatbot use, we found that a large number of users (10,371/16,519, 62.78%) attempted to use the application only once. In some sessions, users presented more than one health issue and never used the application again. For those who used the chatbot more often (n=6148), some used the consultation service on one or two specific days. For example, one user who initiated 23 consultations with the chatbot only used the application on February 20, 2019, and March 13, 2019, to present two different health concerns (headache and palpitation). These findings suggest that even though DoctorBot attracted a lot of users, it was not used extensively. Furthermore, the significant number of "single-use" sessions reveals the issue of user engagement and retention; therefore, it is vital to understand what caused the low usage and how to optimize user experience to achieve the best uptake [23].

What Health Concerns Were Presented to DoctorBot?

Each completed consultation session immediately returned a diagnostic report to the user. A typical report contains the major health concerns expressed by the user and their association with possible diseases that are automatically classified by DoctorBot based on a widely adopted disease classification schema [36]. We analyzed the generated diagnostic reports to understand what health concerns or diseases people usually presented to DoctorBot and the frequency with which each disease was presented. This step was followed by a comparison with the usage of health services in hospitals. In particular, we measured the frequency of consultation on DoctorBot for each disease type and compared that with the health service usage reported by three primary hospitals in China [37]. As Figure 4 shows, diseases with mild symptoms, such as those in the gastroenterology and dermatology categories, appeared in a lot of chatbot consultations, in proportions that were significantly higher than those seen in primary hospitals. One possible explanation is that people with mild symptoms would prefer using the chatbot to query the necessity of clinical visits first, rather than going to hospitals directly. In contrast, using DoctorBot to address emergency care issues was not very common. We also noticed the use of DoctorBot to seek help on medical conditions that often entail considerable privacy and social stigma issues, such as sexually transmitted diseases.

Figure 4. Comparison of the frequency of consultations according to the major disease category of the presenting illness between DoctorBot (orange bars) and primary hospitals (blue bars).



Issues in the Use of DoctorBot

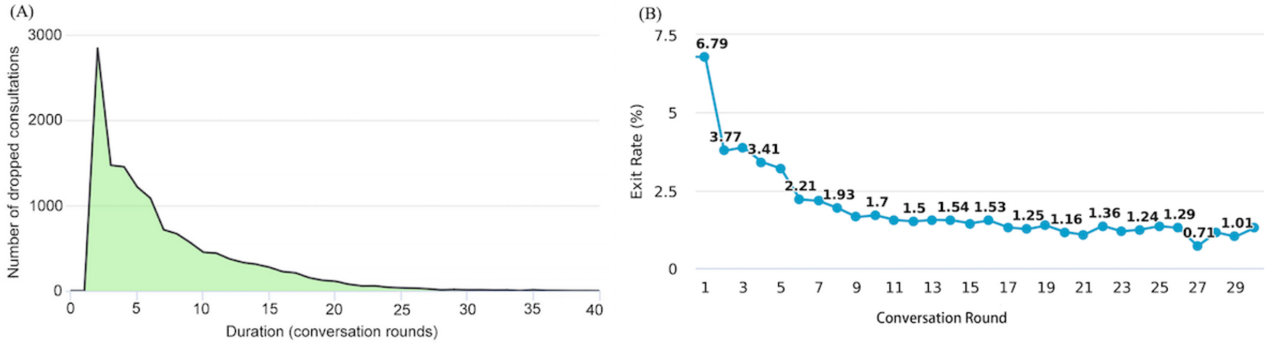
In this section, we report our analysis on the completed and uncompleted sessions, as well as on the user feedback, to present the major issues in the real-world use of DoctorBot.

Dropping Out of Consultations

Our data showed that there was a large number of uncompleted consultations, where users withdrew at a certain point without completing the process (16,974/47,684, 35.60%). In particular, more than one-half of the uncompleted consultations (9542/16,974, 56.22%) occurred within the first five conversation rounds (Figure 5A). To further verify our observation, we built a model to measure the exit rate for each conversation round:

where ER(x) is the exit rate of dropping a conversation in round x, D_x is the number of conversations that drop in round x, and N_x is the number of all conversations.

Figure 5. (A) Distribution of dropped consultation sessions over conversation rounds. (B) The exit rate for each conversation round.



As Figure 5B shows, the exit rate spiked at the beginning, signaling that a lot of users dropped out after a very brief interaction with the chatbot (ie, after just one round of conversation). Furthermore, the exit rate of the first five conservation rounds was much higher than for the rest. These findings are consistent with our observation that users most likely terminated the consultation at an early stage.

To further understand what caused user dropout, we examined the specific question asked by DoctorBot when a consultation was terminated and how much time users spent on answering those questions. As Table 1 demonstrates, transition questions, such as the ones prompting the user to answer the question again, often led to a termination of chatbot use. Furthermore, many uncompleted conversations occurred when users were asked to provide a detailed account of the symptoms they were experiencing. Our analysis also showed that such questions usually took longer than other relatively simple questions, such as questions about demographic information (91.1 s versus 17.6 s, respectively). One possible explanation is that symptom-related questions were usually hard to answer and could easily overwhelm or even confuse users, leaving them unsure of what input to provide and eventually causing them to terminate the conversation. For example, during a consultation for a headache, the chatbot asked whether the user was experiencing a series of symptoms related to headaches, such as fever, vomiting, stuffy nose, cough, and chest tightness. When the chatbot attempted to ask another follow-up question, the user suddenly terminated the consultation.

Table 1. The categories of questions asked by DoctorBot before users terminated the consultation.

Question category	Number of times users terminated their consultation when question was asked (%[a])
Demographic information	58 (4.39)
Physiological data	208 (2.58)
Transition questions	5274 (14.74)
Symptoms	8469 (20.23)
Medical history	795 (4.26)

[a] Calculated as the number of times users terminated their consultation when a question was asked divided by the number of conversations that contained questions from that category.

Lastly, we analyzed the differences between completed and uncompleted sessions to identify the major characteristics of dropped consultations. In particular, we conducted PCA and built a binary logistic regression model using the following features: gender, age, duration of the consultation, number of conversation rounds, and average duration of each round. We included gender and age in the model because these two factors may influence users' acceptance and use of technology [38,39].

As Figure 6 shows, despite that there were twice as many completed consultations as uncompleted consultations, most of the completed consultations (blue dots) were plotted in a small area in the two-dimensional feature space while the dropped consultations (red dots) were distributed horizontally, illustrating that considerable differences existed between completed and uncompleted consultations.

When we controlled the other features (eg, gender, age, and duration of consultation), logistic regression analysis showed that the number of conversation rounds ($P<0.001$) and the average duration of each round ($P<0.001$) significantly correlated with the dependent variable (whether the user dropped out in the middle of the consultation session) (Table 2). More specifically, as the number of conversation rounds with the chatbot increased, the likelihood of terminating the consultation decreased (odds ratio 0.7320, 95% CI 0.7298-0.7341). This finding is consistent with our observation that users most likely terminated the consultation at an early stage.

Figure 6. A principal component analysis (PCA) scatterplot of consultations for 30,710 completed (blue dots) and 16,974 dropped (red dots) consultations. PCA has successfully found linear combinations of the different features in a two-dimensional feature space that separate two different clusters corresponding to whether or not the consultations were completed.

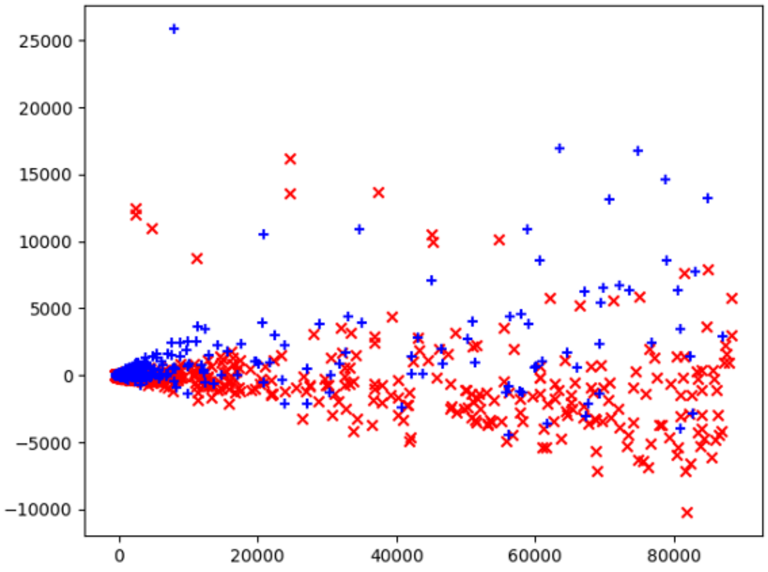


Table 2. The estimated coefficients, odds ratio, and *P* value of each feature in the logistic regression model with terminating consultation as the dependent variable.[a]

Feature	Estimated coefficients (SE)	Odds ratio (95% CI)	*P* value
Gender	3.19e–02 (2.95e–02)	1.0324 (1.0023-1.0633)	.28
Age	1.20e–03 (1.07e–03)	1.0012 (1.0001-1.0023)	.26
Duration of consultation	5.04e–05 (6.64e–06)	1.0001 (1.0000-1.0001)	<.001*
Number of conversation rounds	–3.12e–01 (2.99e–03)	0.7320 (0.7298-0.7341)	<.001*
Mean duration of each conversation round	–3.17e–04 (8.57e–05)	0.9997 (0.9996-0.9998)	<.001*

[a] McFadden pseudo R^2=0.4338884.

*P<.001.

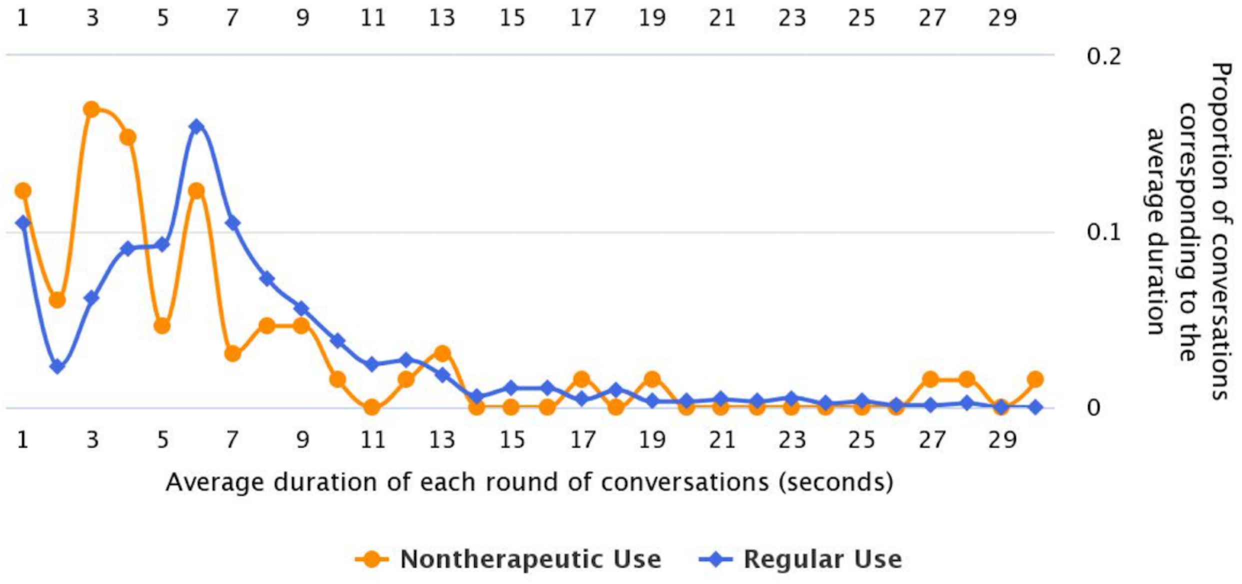
Using Chatbot for Nontherapeutic Purposes

Through the analysis of 3000 randomly selected completed consultation sessions, we noticed another issue in the use of DoctorBot: users often pretended to have health concerns and did not always use the chatbot for therapeutic purposes (ie, medical consultations). In fact, they "gamed" the chatbot, a user behavior that has been reported in intelligent tutoring systems [40] and workplace chatbot systems [41]. This behavior was exhibited in 241 of 3000 (8.03%) consultation sessions. We characterized the "nontherapeutic" use of DoctorBot into 5 patterns (Table 3). For example, people used nonsense/illogical words (eg, "I miss you") or contradictory statements to describe their health concerns. They might have even used the same word (eg, "nothing," "unclear") throughout the entire consultation. We also compared regular and nontherapeutic consultation sessions in terms of the mean duration of conversations. It was shown that the proportion of nontherapeutic and regular sessions with a conversation round lasting up to 4 s was 50.76% and 17.53%, respectively (Figure 7). This finding revealed that users who used the chatbot for nontherapeutic purposes attempted to complete each round of conversation as quickly as possible. We also analyzed the standard deviation of time taken per conversation round, which showed that users with real health concerns spent various lengths of time on different questions (mean 11.85 s, SD 17.06 s). In contrast, users who didn't actually use the chatbot for medical consultation spent an equally short time on each question (mean 7.74 s, SD 14.33 s).

Table 3. Patterns of nontherapeutic chatbot use and illustrative examples.

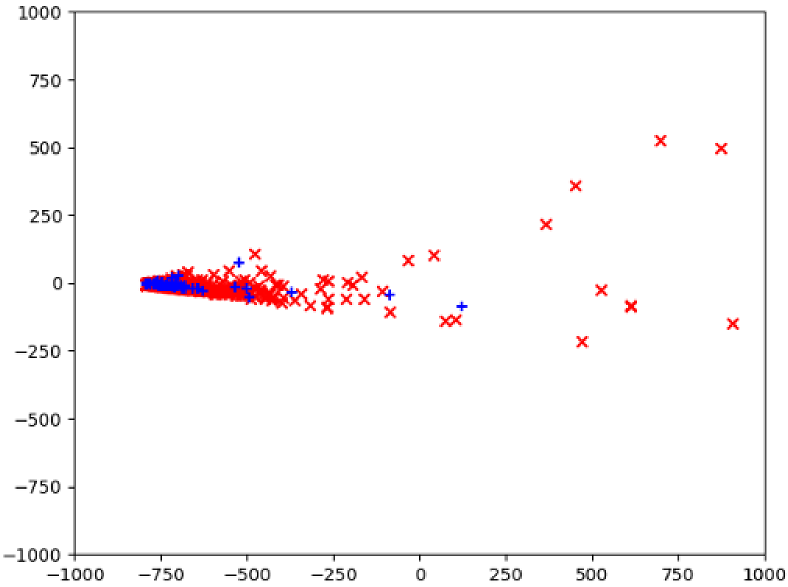
Patterns	Example
Extensive use of nonsense/illogical words	Chabot: If you have any other symptoms, please continue typing. User: I miss you.
Contradictory statements	Chatbot: Please tell me your symptoms. User: How can I deal with a headache? Chatbot: Which of the following symptoms does your description fit? User: Fever. Chatbot: What is your highest temperature? User: Unknown. Chatbot: How about the degree of headache? User: Painless.
Continuously use the same word to answer all the questions	Chatbot: Please tell me your symptoms. User: Headache. Chatbot: What kind of headache is it? User: Unclear. Chatbot: Do you have dizziness? User: Unclear. Chatbot: Is it accompanied by elevated blood pressure, chills, and other symptoms? User: Unclear. Chatbot: Is it accompanied by neck pain? User: Unclear. […]
Answering each question quickly	In the previous example, the user went through 19 rounds and spent 49 seconds, averaging only 2.6 seconds per question.
Spending an equally short time on each question	In the previous example, the user answered almost all of the questions in 2-3 seconds, and the SD of time was 1.2 seconds.

Figure 7. The distribution of the average duration of conversations.



To further verify our findings of the patterns of nontherapeutic chatbot use, we conducted PCA and built a binary logistic regression model using the following features: gender, age, duration of a consultation, number of conversation rounds, mean duration of each round, time spent in each round, and number of conversation rounds with continuous repetition of the same response. We did not consider contradictory statements or extensive use of nonsense/illogical words in the analysis because it was not appropriate to quantify them. As shown in Figure 8, the regular consultations (blue dots) are plotted in a small area (the number of blue dots is 10 times greater than the number of red dots), whereas the nontherapeutic use cases (red dots) are scattered. This shows that there were obvious abnormalities in many nontherapeutic use cases.

Figure 8. A principal component analysis (PCA) scatterplot of consultations for 2759 regular (blue dots) and 241 nontherapeutic (red dots) consultations. PCA has successfully found linear combinations of the different features in a two-dimensional feature space that separates two different clusters corresponding to whether or not the chatbot was used for a therapeutic purpose.



Our logistic regression analysis showed that, when controlling other features, the mean duration of each round (*P*=0.03), the standard deviation of time spend in each round (*P*=0.05), and repetition of the response (*P*<0.001) significantly correlated with the occurrence of nontherapeutic chatbot use (Table 4). This suggests that for a consultation session, the shorter the average duration of each round was, the more likely the user was gaming the chatbot (odds ratio 0.9618, 95% CI 0.9409-0.9892). Also, the result suggests that if the duration of different conversation rounds was more varied, users were less likely to be using the chatbot for a nontherapeutic purpose (odds ratio 0.9835, 95% CI 0.9733-0.9939).

Table 4. The estimated coefficients, odds ratio, and *P* value of each feature in the logistic regression model, with using the chatbot for nontherapeutic purposes as the dependent variable.[a]

Feature	Estimated coefficients (SE)	Odds ratio (95% CI)	*P* value
Gender	6.28e–02 (2.98e–02)	1.0648 (1.0335-1.0970)	.09
Age	–1.66e–02 (1.05e–02)	0.9835 (0.9733-0.9939)	.11
Duration of consultation	–3.31e–03 (1.81e–03)	0.9967 (0.9949-0.9985)	.07
Number of conversation rounds	–4.60e–02 (3.17e–02)	0.9550 (0.9252-0.9858)	.15
Average duration of each round	–3.89e–02 (2.20e–02)	0.9618 (0.9409-0.9892)	.03*
SD of time spent in each round	–2.31e–02 (1.22e–02)	0.9772 (0.9653-0.9892)	.05*
Rounds of continuous repetition of the same response (unclear or unknown)	4.56e–01 (6.24e–02)	1.5778 (1.4823-1.6793)	<.001**

[a] McFadden pseudo R^2=0.2478434.

*$P<.05$.

**$P<.001$.

User Concerns

Toward the end of each consultation, DoctorBot prompted the user to rate the experience as either positive or negative (Figure 9). If a negative rating was chosen, the system asked the user to provide further feedback, which users could choose to do by either typing into a comment box or selecting from a list of predefined reasons (eg, “diagnosis is not accurate” or “overwhelming information”). Despite this step being optional, many users provided feedback. In total, we collected 3832 pieces of user feedback, with 2172 positive ratings and 1660 negative ratings. As negative ratings usually suggest that users had concerns, we examined the specific reasons for negative ratings by analyzing the textual feedback and the predefined reasons selected by users.

Figure 9. Interface for collecting user feedback: rating (left), predefined reasons (top right), and comment box (bottom right).

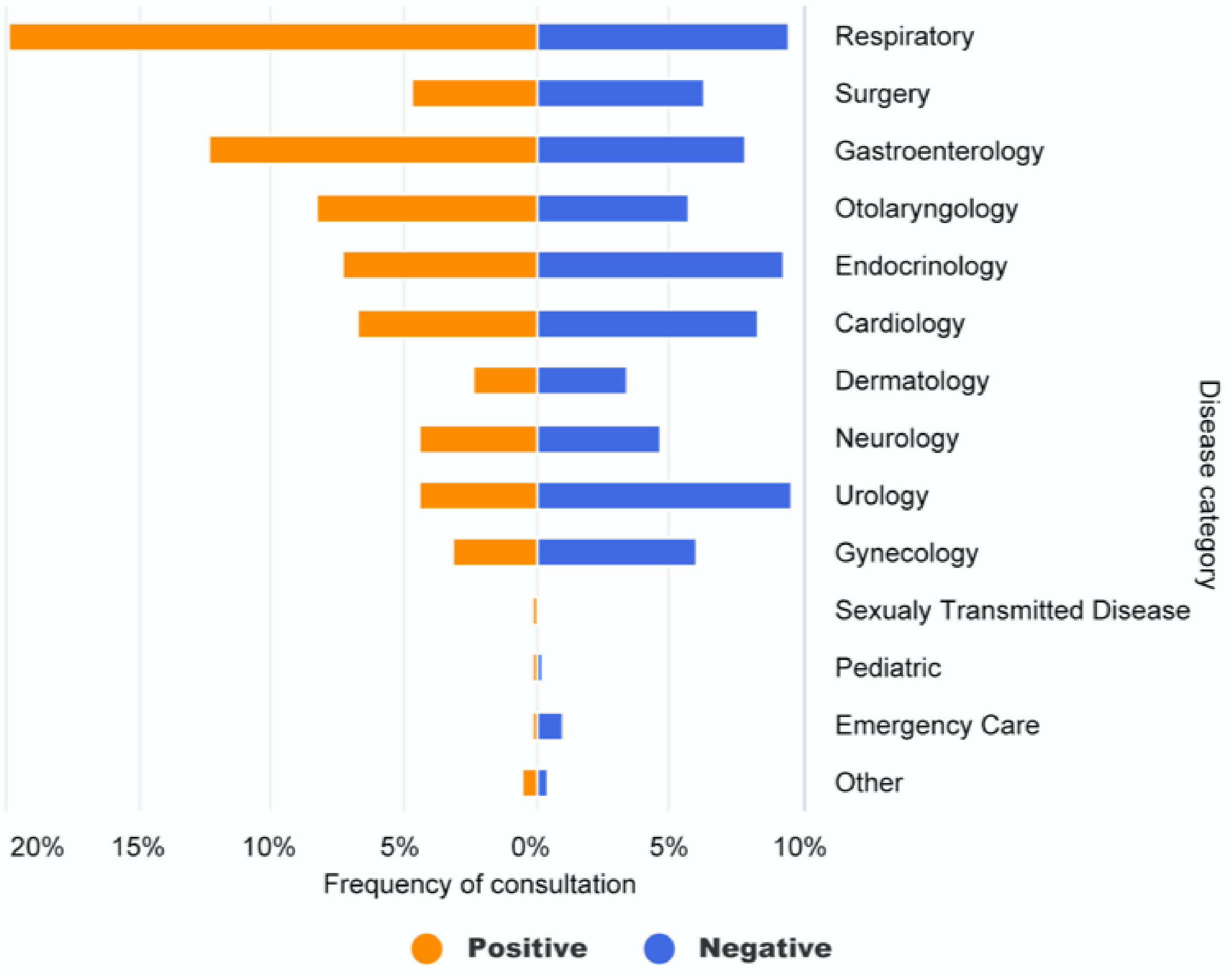
We characterized five user concerns that led to negative ratings (Table 5). We found that users usually expressed dissatisfaction when they believed that the diagnostic suggestion was not accurate; as one user stated, “I’m 21 years old with a regular lifestyle. I don’t smoke or drink. I eat and exercise regularly. You told me I have diabetes!!!” Some users even compared the diagnostic suggestions of DoctorBot with their physician’s diagnosis and stated that there was a discrepancy: “My doctor told me there is nothing to worry about and prescribed me an herbal medication. But the chatbot suggested a different diagnosis and I don’t trust it.” Furthermore, if users had difficulty comprehending the diagnostic report generated by DoctorBot or determining which diagnostic suggestion was more reliable, users also gave negative ratings. For example, one user commented, “I have no clue about the suggested diagnosis. What is that?” In other cases, users complained that the provided information on the diagnostic report was overwhelming. Indeed, due to different levels of health literacy, knowledge, and experience, users may have had challenges in comprehending the technical aspects of the diagnostic report (eg, medical jargon and professional medical knowledge) [42-44]. Lastly, users tended to give negative ratings if their information needs were not fully met. The analysis of textual feedback revealed that users desired to receive more personalized and actionable information, including medical information related to their health concerns, where to seek medical help, what to do next, and detailed explanations about the suggested diagnoses.

Table 5. Reasons cited by DoctorBot users for giving a negative rating of their experience.

User concern	Number of users
Suggested diagnosis is perceived to be inaccurate	1084
Difficult to assess, which suggested diagnosis is correct	164
Insufficient information	247
Report is not easy to understand	113
Provided information is overwhelming	52

We also analyzed the relationship between negative ratings and several intrinsic factors (eg, age, gender, duration of consultation, and disease type) to further examine if the user experience was affected by other factors that were not revealed by the analysis of user feedback. In particular, we found that the total time spent on the consultation had a significant impact on user experience. If a consultation lasted more than 2 minutes, users tended to rate their experience as negative. Interestingly, the disease type was also highly related to users’ experiences and their satisfaction level (Figure 10). For example, medical advice about common diseases, such as respiratory issues, usually received positive ratings. One possible explanation is that the chatbot could easily diagnose these diseases and provide pertinent information and medical advice to fulfill the users’ needs. However, it is challenging for the chatbot to provide an accurate diagnosis of and meaningful information about diseases with complex causal mechanisms, such as gynecopathy, based on only a few rounds of conversation; as such, negative ratings against DoctorBot-generated outputs were fairly common under such circumstances.

Figure 10. Relationship between user ratings and disease category of presenting illness of DoctorBot consultations.



Discussion

Principal Findings

The scarceness and imbalanced distribution of health care resources (eg, facilities and doctors) are major health concerns worldwide [7,8]. Many people, especially those in rural areas, may not have immediate and convenient access to the medical services they need. Furthermore, due to the complex organization and workflow of conventional health care services, it is challenging for patients and caregivers to navigate the health care system [45]. As such, more digital tools have been introduced to help patients triage before seeing a doctor and gain information to fill their knowledge gap. For example, as predecessors of self-diagnosis chatbots, online symptom checkers have been launched to more effectively provide possible alternative diagnoses to patients and direct them to the appropriate care settings [46]. Prior work illustrated that symptom checkers have an acceptable level of patient compliance with medical advice [47] and triage accuracy [48]. However, symptom checkers often use some proprietary diagnostic algorithms, such as branching logic and Bayesian inference, that are not optimal for processing complicated and sometimes ambiguous user inputs, compromising not only user experience but also diagnostic accuracy [47,49].

With the recent advent of AI technologies such as knowledge graphs and deep neural networks, more sophisticated self-diagnosis health chatbots are emerging to simulate the conversations between patient and care providers and provide more accurate and comprehensive care advice and services. The application of chatbot technology in the health domain, including mental health [21] and behavioral therapy [50,51], is becoming more and more common. Despite the effectiveness of chatbots in delivering health care services to improve well-being, this novel technology has not been adopted at the rates predicted based on the high level of interest [25]. Due to the lack of large-scale deployment, prior work primarily examined the use of health chatbots in controlled settings rather than in a real-world context [28]. This study aimed to bridge the knowledge gap in how people use health chatbots in real-world scenarios.

To that end, we analyzed the system log of a self-diagnosis chatbot to understand how health care chatbots would be used in the real world and what issues could impede the optimal use of this novel technology. We found that users in all age ranges, including middle-aged and older adults, had used the chatbot. A considerable number of people used the chatbot only once. Users consulted the chatbot about a wide range of topics, including mild medical conditions, as well as those that often entail considerable privacy and social stigma issues. We also observed several issues in the use of the chatbot, including user dropout and use for nontherapeutic purposes. Finally, we identified a set of user concerns that should be addressed to optimize user experience, including receiving insufficient actionable information and perceived inaccurate diagnostic suggestions. We argue that designers and developers of health chatbots need to employ user-centered approaches to address users' concerns and issues. Below we discuss design implications for health chatbots to enhance user experience and engagement.

Design Implications

Designing Informative Health Chatbots

The analysis of user feedback revealed that users expressed the need to receive more actionable information, such as next steps to take. This suggests that users' information needs were not adequately met by the chatbot. Also, users complained that the system-generated diagnostic report was difficult to interpret. These findings highlight the importance of providing more useful information that patients need. For example, in the diagnostic report, chatbots could provide links to consumer-friendly and credible information sources to help patients better understand the content of the report [42].

Designing Easy-To-Use Health Chatbots

Chatbots typically ask consecutive questions about concomitant symptoms so that they can generate a more accurate diagnosis. However, these questions are usually hard to answer and can easily overwhelm the user. For example, we found that users tended to terminate the consultation when they were asked to describe their symptoms or chief complaints. To address these issues, it would be useful to allow users to share and describe information in the form of voice recordings to reduce the amount of time and effort spent on typing. The chatbots should also be designed to inform users as to why a particular piece of information is needed [52].

Another interesting finding is that a small number of elderly people also used the application. Given the rise in the aging population and its associated health care costs in many countries, health chatbots will likely become a promising approach to aid older adults' independent living [53]. This raises a new set of questions for chatbot designers about how to make health chatbots more accessible and user-friendly for older adults. For example, a speech interface can be incorporated into chatbot systems to facilitate communication between elderly users and health chatbots [54].

Designing Trustworthy Health Chatbots

Through the analysis of user feedback, we found that perceiving the chatbot's output (eg, recommended diagnosis) as inaccurate was highly associated with a negative user experience. This is not surprising because health care has a high degree of criticality and users tend to have doubts about the diagnosis suggested by chatbots. Therefore, to better engage users, it is very important to increase the trustworthiness of health chatbots. Prior literature on AI-driven intelligent systems has suggested presenting system outputs in a format that is meaningful, understandable, and trustworthy to help users better understand the system's hidden intelligence and then determine whether it is appropriate to trust the recommendations and use them to make decisions [55]. In particular, prior work suggested presenting a variety of system-related information to the user, including system reliability and performance data, logic, and reasoning (eg, how the system operates and how its outputs are generated), as well as the information sources that the system leverages to produce the output [52,56,57]. Aligning with those arguments, we suggest that health chatbots need to explain data sources, prediction accuracy, and how the diagnostic report is generated to the users to build trust. For example, chatbots could provide more appropriate explanations to indicate what types of diseases they are knowledgeable of and their degree of confidence in their diagnostic suggestion.

Designing Onboarding Experiences for Users

We observed that many users dropped out of consultations, especially during their early phases. This finding highlights the necessity of enhancing user engagement at an early point. Furthermore, the chatbot was sometimes used for nontherapeutic purposes. We speculate that because self-diagnosis chatbots are an emerging technology, some users may just want to navigate through the application to explore how the chatbot works. However, gaming the chatbot could generate a large amount of noisy data, some of which might be used to train models; therefore, nontherapeutic use cases, if not taken care of properly, could adversely affect the performance of health chatbots. To prevent these issues and better engage users, it may be useful to provide them with onboarding materials during the initial interactions. Prior work has suggested that onboarding materials could educate users about the most effective way to use advanced technologies (ie, AI-driven health chatbots) [58]. As an example, the onboarding materials could introduce users to the basic functions (eg, capabilities and limitations) of the chatbot and the process of consultation (eg, what types of questions will be asked and why). Moreover, the chatbot should be designed to automatically detect and tag nontherapeutic use cases so that developers could easily remove such noisy data when training AI models.

Limitations

The present study has several limitations. First, this study relied heavily on log data. Even though the analysis of log data could provide valuable insights into the use of the chatbot in real-world settings, it didn't allow us to capture the perceptions and opinions of users when interacting with the chatbot, such as what features they liked and disliked, what barriers they encountered, and how the chatbot should be improved to optimize the user experience. In the future, we will employ social theories on the explanatory factors of the use of technology (eg, diffusion of innovation) to conduct user studies (eg, interview, survey, and usability evaluations) with different groups of people to form a more comprehensive view of users' attitudes and experiences about this novel technology. Second, we only examined the use of one health chatbot, which is likely to compromise the generalizability of the findings. To assess and expand our results' generalizability, it would be useful to examine other chatbots, including specialist chatbots that serve a particular population with specific conditions. Lastly, cultural and social factors could also play a vital role in the utilization of health chatbots. We will also examine those issues in our future work.

Conclusions

In this paper, we conducted both quantitative and qualitative analysis on the system log of a self-diagnosis chatbot, which has been widely deployed in China. To the best of our knowledge, this is the first study examining the use of health chatbots in the real world using a large-scale, heterogeneous data set. We described our general observations of the chatbot's use, including who used it, how long and how often they used the application, and what health concerns were presented. Furthermore, we analyzed both completed and uncompleted consultation sessions as well as the user feedback to investigate issues that may hinder the effective use of the chatbot. These results shed light on the design of health chatbots to improve user experience and increase user engagement.

Acknowledgments

This work was supported by the National Key RD Program of China (Grant No. 2017YFB1002504), Beijing NOVA Program (Grant No. Z201100006820120), and the CAS Pioneer Hundred Talents Program.

Conflicts of Interest

None declared.

References

1. Ma X, Gui X, Fan J, Zhao M, Chen Y, Zheng K. Professional Medical Advice at your Fingertips. In: Proc. ACM Hum.-Comput. Interact. 2018 Nov Presented at: Proceedings of the ACM on Human-Computer Interaction, 2(CSCW); November 3rd-7th; Jersey City p. 1-22 URL: https://doi.org/10.1145/3274385 [doi: 10.1145/3274385]
2. Gui X, Chen Y, Kou Y, Pine K, Chen Y. Investigating Support Seeking from Peers for Pregnancy in Online Health Communities. In: Proc. ACM Hum.-Comput. Interact. 2017 Dec 06 Presented at: Proceedings of the ACM on Human-Computer Interaction, 1(CSCW); February 25-March 1; Portland p. 1-19 URL: https://doi.org/10.1145/3134685 [doi: 10.1145/3134685]
3. Kuehn BM. Global shortage of health workers, brain drain stress developing countries. JAMA 2007 Oct 24;298(16):1853-1855. [doi: 10.1001/jama.298.16.1853] [Medline: 17954532]
4. Laumer S, Maier CF, Gubler FT. Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis. 2019 Presented at: Proceeding of the 27th European Conference on Information System (ECIS); June 8-14; Stockholm&Uppsala, Sweden p. E URL: https://aisel.aisnet.org/ecis2019_rp/88 [doi: 10.1145/3322385.3322392]
5. Ensor T, Cooper S. Overcoming barriers to health service access: influencing the demand side. Health Policy Plan 2004 Mar;19(2):69-79. [Medline: 14982885]
6. Gui XN, Chen YN. Making healthcare infrastructure work: Unpacking the infrastructuring work of individuals. 2019 Presented at: Proceedings of the ACM Conference on Human Factors in Computing Systems; May 4-7; Glasgow, United Kingdom p. 1-14 URL: https://doi.org/10.1145/3290605.3300688 [doi: 10.1145/3290605.3300688]
7. Chen YN, Cheng K, Tang C, Siek KA, Bardram JE. Is my doctor listening to me? Impact of health IT systems on patient-provider interaction. 2013 Presented at: CHI'13 Extended Abstracts on Human Factors in Computing Systems; April 27-May 2; Paris, France p. 2419-2426 URL: https://doi.org/10.1145/2468356.2468791 [doi: 10.1145/2468356.2468791]
8. Park SY, Chen YN, Raj S. Beyond health literacy: Supporting patient-provider communication during an emergency visit. 2017 Presented at: Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing; February 25-March 1; Portland p. 2179-2192 URL: https://doi.org/10.1145/2998181.2998357 [doi: 10.1145/2998181.2998357]
9. Björk AB, Hillborg H, Augutis M, Umefjord G. Evolving techniques in text-based medical consultation - Physicians' long-term experiences at an Ask the doctor service. Int J Med Inform 2017 Sep;105:83-88 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.05.013] [Medline: 28750914]
10. Umefjord G, Sandström H, Malker H, Petersson G. Medical text-based consultations on the Internet: a 4-year study. Int J Med Inform 2008 Feb;77(2):114-121. [doi: 10.1016/j.ijmedinf.2007.01.009] [Medline: 17317292]
11. Vinker S, Weinfass M, Kasinetz LM, Kitai E, Kaiserman I. Web-based question-answering service of a family physician – the characteristics of queries in a non-commercial open forum. Med Inform Internet Med 2007 Jun;32(2):123-129. [doi: 10.1080/14639230601178653] [Medline: 17541862]
12. Zhang Z, Lu Y, Kou Y, Wu DTY, Huh-Yoo J, He Z. Stud Health Technol Inform 2019 Aug 21;264:1403-1407 [FREE Full text] [doi: 10.3233/SHTI190458] [Medline: 31438157]
13. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
14. Pemberton PJ, Goldblatt J. The Internet and the changing roles of doctors, patients and families. Med J Aust 1998;169(11-12):594-595. [Medline: 9887903]
15. Swire-Thompson B, Lazer D. Public health and online misinformation: Challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41:433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]

16. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
17. Iacobucci G. NHS long term plan: Care to be shifted away from hospitals in "21st century" service model. BMJ 2019 Jan 07;364:l85. [doi: 10.1136/bmj.l85] [Medline: 30617145]
18. Curioni-Fontecedro A. A new era of oncology through artificial intelligence. ESMO Open 2017;2(2):e000198 [FREE Full text] [doi: 10.1136/esmoopen-2017-000198] [Medline: 28761754]
19. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009 May;46(1):5-17 [FREE Full text] [doi: 10.1016/j.artmed.2008.07.017] [Medline: 18790621]
20. Ivanovic M, Semnic M. The role of agent technologies in personalized medicine. 2018 Presented at: 5th International Conference on Systems and Informatics (ICSAI); November 10-12; Nanjing, China p. 299-304 URL: https://doi.org/10.1109/ICSAI.2018.8599421 [doi: 10.1109/icsai.2018.8599421]
21. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
22. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. 2017 Presented at: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May; Barcelona, Spain p. 261-265 URL: https://doi.org/10.1145/3154862.3154914 [doi: 10.1145/3154862.3154914]
23. Montenegro J, da Costa C, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
24. Tripathy AK, Carvalho R, Pawaskar K, Yadav S, Yadav V. Mobile based healthcare management using artificial intelligence. 2015 Presented at: International Conference on Technologies for Sustainable Development (ICTSD); Feb 4-6; Mumbai, India p. 1-6 URL: https://doi.org/10.1109/ICTSD.2015.7095895 [doi: 10.1109/ictsd.2015.7095895]
25. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
26. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 1966 Jan;9(1):36-45 [FREE Full text] [doi: 10.1145/365153.365168]
27. Madhu D, Jain C, Sebastain E, Shaji S, Ajayakumar A. A novel approach for medical assistance using trained chatbot. 2017 Presented at: International Conference on Inventive Communication and Computational Technologies; March 10-11; Coimbatore, India p. 243-246 URL: https://doi.org/10.1109/ICICCT.2017.7975195 [doi: 10.1109/icicct.2017.7975195]
28. Pereira J, Díaz Ó. Using Health Chatbots for Behavior Change: A Mapping Study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
29. Zuo Shou Yi Sheng. URL: https://www.zuoshouyisheng.com/ [accessed 2020-12-04]
30. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a Chatbot (Ada) in the Diagnosis of Mental Disorders: Comparative Case Study With Lay and Expert Users. JMIR Form Res 2019 Oct 29;3(4):e13863 [FREE Full text] [doi: 10.2196/13863] [Medline: 31663858]
31. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. arXiv Preprint posted online June 27, 2018. [FREE Full text]
32. Doshi-Velez F, Kortz M, Budish R, Bavitz C, Gershman S, O'Brien D, et al. Accountability of AI Under the Law: The Role of Explanation. SSRN preprints Preprint posted online December 22, 2019. [doi: 10.2139/ssrn.3064761]
33. Jieba. URL: https://pypi.org/project/jieba/ [accessed 2020-12-04]
34. Mayring P. Qualitative content analysis. A Companion to Qualitative Research, 1, 159-176. ISBN, 978076197 2004:3744. [doi: 10.1093/oxfordhb/9780199811755.013.008]
35. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
36. ICD-11, International Classification of Diseases 11th Revision. URL: https://icd.who.int/ [accessed 2020-12-04]
37. Aihua L, Qing L. An analysis of disease classification of patients in different levels of hospitals in Guangdong. Chinese Journal of Hospital Statistics 2000;3:n/a.
38. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Information & Management 2003 Jan;40(3):191-204 [FREE Full text] [doi: 10.1016/s0378-7206(01)00143-4]
39. Zarouali B, Van den Broeck E, Walrave M, Poels K. Predicting Consumer Responses to a Chatbot on Facebook. Cyberpsychol Behav Soc Netw 2018 Aug;21(8):491-497. [doi: 10.1089/cyber.2017.0518] [Medline: 30036074]
40. Baker R, Walonoski J, Heffernan N, Roll I, Corbett A, Koedinger K. Why students engage in "gaming the system" behavior in interactive learning environments. Journal of Interactive Learning Research 2008;19(2):185.
41. Liao Q, Mas-ud Hussain M, Chandar P, Davis M, Khazaeni Y, Crasso M, et al. All work and no play? 2018 Presented at: Proceedings of the 2018 ACM Conference on Human Factors in Computing Systems; April 21-26; Montreal, Canada p. 1-13 URL: https://doi.org/10.1145/3173574.3173577 [doi: 10.1145/3173574.3173577]

42. Hong MK, Feustel C, Agnihotri M, Silverman M, Simoneaux SF, Wilcox L. Supporting families in reviewing and communicating about radiology imaging studies. 2017 Presented at: Proceedings of the 2017 ACM Conference on Human Factors in Computing Systems; May 6-11; Denver. [doi: 10.1145/3025453.3025754]
43. Reynolds TL, Ali N, McGregor E, O'Brien T, Longhurst C, Rosenberg AL, et al. Understanding patient questions about their medical records in an online health forum: Opportunity for patient portal design. 2017 Presented at: AMIA Annu Symp Proc; Nov 4-8; Washington, D.C p. 1468-1477 URL: http://europepmc.org/abstract/MED/29854216
44. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
45. Bodenheimer T. Coordinating care-a perilous journey through the health care system. N Engl J Med 2008 Mar 6;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]
46. North F, Ward WJ, Varkey P, Tulledge-Scheitel SM. Should you search the Internet for information about your acute symptom? Telemed J E Health 2012 Apr;18(3):213-218. [doi: 10.1089/tmj.2011.0127] [Medline: 22364307]
47. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
48. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015;351:h3480 [FREE Full text] [Medline: 26157077]
49. Van Allen EM, Wagle N, Levy MA. Clinical analysis and interpretation of cancer genome data. J Clin Oncol 2013 May 20;31(15):1825-1833 [FREE Full text] [doi: 10.1200/JCO.2013.48.7215] [Medline: 23589549]
50. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
51. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M, et al. Pocket skills: A conversational mobile web app to support dialectical behavioral therapy. 2018 Presented at: Proceedings of the ACM Conference on Human Factors in Computing Systems; April 21-26; Montreal, Canada p. 1-15 URL: https://doi.org/10.1145/3173574.3173972 [doi: 10.1145/3173574.3173972]
52. Vorm ES. Assessing demand for transparency in intelligent systems using machine learning. 2018 Presented at: Innovations in Intelligent Systems and Applications (INISTA); July 3-5; Thessaloniki, Greece p. 1-7 URL: https://doi.org/10.1109/INISTA.2018.8466328 [doi: 10.1109/insta.2018.8466328]
53. Ghosh S, Bhatia S, Bhatia A. Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System. Stud Health Technol Inform 2018;252:51-56. [Medline: 30040682]
54. Blair J, Abdullah S. Understanding the needs and challenges of using conversational agents for deaf older adults. 2019 Presented at: Conference Companion Publication on Computer Supported Cooperative Work and Social Computing; Nov 9-13; Austin p. 161-165 URL: https://doi.org/10.1145/3311957.3359487 [doi: 10.1145/3311957.3359487]
55. Vorm ES, Miller AD. Assessing the value of transparency in recommender systems: An end-user perspective. 2018 Presented at: Proceedings of the 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with ACM Conference on Recommender Systems; October 2-7; Vancouver, Canada p. 61-68 URL: http://ceur-ws.org/Vol-2225/ [doi: 10.1145/2645710.2645787]
56. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
57. Fisher CW, Kingma BR. Criticality of data quality as exemplified in two disasters. Information & Management 2001 Dec;39(2):109-116 [FREE Full text] [doi: 10.1016/s0378-7206(01)00083-0]
58. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. 2019 Nov 07 Presented at: Proceedings of the ACM on Human-Computer Interaction, 3(CSCW); Nov 9-13; Austin p. 1-24 URL: https://doi.org/10.1145/3359206 [doi: 10.1145/3359206]

Abbreviations

AI: artificial intelligence
mHealth: mobile health
PCA: principal component analysis

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by E Bellei, I Gabashvili, J Torrent Sellens; comments to author 28.05.20; revised version received 21.07.20; accepted 02.10.20; published 06.01.21.

Please cite as:
Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F
Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study
J Med Internet Res 2021;23(1):e19928
URL: https://www.jmir.org/2021/1/e19928
doi:10.2196/19928
PMID:33404508

Review

Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review

Alaa A Abd-Alrazaq[1], BSc, MSc, PhD; Mohannad Alajlani[2], BSc, MSc, MBA, PhD; Nashva Ali[1], BSc, MSc; Kerstin Denecke[3], Dr rer nat; Bridgette M Bewick[4], BA, MA, PhD; Mowafa Househ[1], BCom, MEng, PhD

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[2]Institute of Digital Healthcare, University of Warwick, Warwick, United Kingdom

[3]Institute for Medical Informatics, Bern University of Applied Science, Bern, Switzerland

[4]Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Mowafa Househ, BCom, MEng, PhD
Division of Information and Computing Technology
College of Science and Engineering, Hamad Bin Khalifa University
Qatar Foundation
PO Box 5825, Doha Al Luqta St, Ar-Rayyan
Doha, RT542
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: Chatbots have been used in the last decade to improve access to mental health care services. Perceptions and opinions of patients influence the adoption of chatbots for health care. Many studies have been conducted to assess the perceptions and opinions of patients about mental health chatbots. To the best of our knowledge, there has been no review of the evidence surrounding perceptions and opinions of patients about mental health chatbots.

Objective: This study aims to conduct a scoping review of the perceptions and opinions of patients about chatbots for mental health.

Methods: The scoping review was carried out in line with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) extension for scoping reviews guidelines. Studies were identified by searching 8 electronic databases (eg, MEDLINE and Embase) in addition to conducting backward and forward reference list checking of the included studies and relevant reviews. In total, 2 reviewers independently selected studies and extracted data from the included studies. Data were synthesized using thematic analysis.

Results: Of 1072 citations retrieved, 37 unique studies were included in the review. The thematic analysis generated 10 themes from the findings of the studies: usefulness, ease of use, responsiveness, understandability, acceptability, attractiveness, trustworthiness, enjoyability, content, and comparisons.

Conclusions: The results demonstrated overall positive perceptions and opinions of patients about chatbots for mental health. Important issues to be addressed in the future are the linguistic capabilities of the chatbots: they have to be able to deal adequately with unexpected user input, provide high-quality responses, and have to show high variability in responses. To be useful for clinical practice, we have to find ways to harmonize chatbot content with individual treatment recommendations, that is, a personalization of chatbot conversations is required.

(J Med Internet Res 2021;23(1):e17828) doi:10.2196/17828

KEYWORDS

chatbots; conversational agents; mental health; mental disorders; perceptions; opinions; mobile phone

Introduction

Background

Mental disorders are a growing global concern. Approximately 29% of individuals may experience such disorders in their lifetime [1]. Mental disorders have a negative effect on the quality of life and are one of the most common causes of disability [2]. Globally, the lost labor and capital output caused by mental disorders are predicted to cost US $16 trillion between 2011 and 2030 [3]. For many, accessing mental health care services is challenging because of the shortage of mental health care providers [4-7], lack of health care insurance coverage [8], and perceived stigma [9-11]. The lack of access to mental health care services increases the risk of self-harm and suicide [12,13].

Technological advancements have improved access to mental health care services [3]. According to the World Health Organization, more than one-fourth of 15,000 mobile health (mHealth) apps focus on mental health diagnosis or support [13]. Chatbots, also called conversational agents, virtual agents, and dialog systems, are one of the main mobile apps used for mental health.

Chatbots are programs able to converse and interact with a human using voice, text, and animation [14]. There are 2 types of chatbots: rule-based chatbots and intelligent chatbots. Although the former uses some predefined rules or decision trees to manage its response and dialog, the latter uses artificial intelligence (AI) to generate its dialog [14]. Chatbots have been employed in businesses across different industries, such as e-commerce and retail (eg, eBay’s ShopBot), hospitality (eg, Chatobook), real estate (eg, Apartment Ocean), entertainment (eg, Mojihunt), news (CNN’s Chatbot), travel (eg, Hello Hipmunk), finance and banking (eg, Wells Fargo’s chatbot), insurance (eg, ABIE), education (eg, Feed.Mind), legal (eg, Immigration Virtual Assistant), and personal assistance (eg, Apple’s Siri). In addition to the abovementioned industries, chatbots have become popular in health care (in general) and mental health (in specific) in the past 5 years. According to a scoping review conducted by Abd-alrazaq et al [14], there were 41 different chatbots used for several purposes in mental health, such as therapy, training, education, counseling, and screening. A systematic review of 12 studies showed that chatbots are effective in improving some mental disorders, such as depression, stress, and acrophobia [15].

Research Problem and Aim

The adoption of new technology relies on the perceptions and opinions of users. Numerous studies have been conducted to assess the perceptions and opinions of patients about mental health chatbots [14]. There is a need to explore perceptions and opinions that may help in predicting the adoption of chatbots and improving them [14]. Although several reviews have been conducted on chatbots in mental health [4,14-17], none have summarized the findings of previous studies regarding perceptions and opinions of patients about mental health chatbots. Accordingly, the aim of this study is to review the perceptions and opinions of patients about mental health chatbots, as reported in the previous literature.

Methods

Study Design

We conducted a scoping review to accomplish this objective. A scoping review was conducted as the aim was to map the body of literature on this topic [18]. Owing to the broad nature of the inquiry, we expected a diversity of study designs. Scoping reviews are more suited to broader aims [18]. As we were not seeking to summarize the best available research on a specific question, a systematic review was not the ideal choice. Using chatbots for mental health is a field in relative infancy; therefore, a broader aim was necessary. The range of study designs currently used in the field makes equitable risk of bias assessment difficult; it is acknowledged that the risk of bias assessments is not required in scoping reviews [18,19]. Scoping reviews are generally accepted as more appropriate when diversity of study designs is expected. The PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Extension for Scoping Reviews guidelines were followed to carry out a systematic and transparent review [20].

Search Strategy

Search Sources

The following electronic databases were searched in the current review: MEDLINE (via Ovid), Embase (via Ovid), PsycINFO (via Ovid), Scopus, Cochrane Central Register of Controlled Trials, IEEE Xplore, ACM Digital Library, and Google Scholar. Given that Google Scholar usually finds several thousands of references, which are ordered by their relevance to the search topic, we screened only the first 100 references [14,15,21]; these references are the most relevant. The search was conducted from October 25 to October 28, 2019. We also conducted backward reference list checking, where reference lists of the included studies and reviews on the search topic were screened for additional studies of relevance to the review. In addition, we carried out forward reference list checking, where the *cited by* function available in Google Scholar was used to find and screen studies that cited the included studies.

Search Terms

To derive search terms, we checked previous literature reviews [4,14-17] and opinions of informatics experts interested in mental health (which were collected in informal meetings). The search terms were selected based on the target intervention (eg, chatbots and conversational agents) and condition (eg, mental disorder and anxiety). The controlled vocabulary search terms (eg, MeSH, Emtree) were used, as appropriate. The search strings used for searching each electronic database are detailed in Multimedia Appendix 1.

Study Eligibility Criteria

The intervention of interest in this review was chatbots that operate as stand-alone software or a web browser (Textboxes 1 and 2). We excluded chatbots that were integrated into robotics, serious games, SMS, or telephone systems and those that depended on human operator–generated dialog. No restrictions were applied regarding the type of dialog initiative (ie, use, system, mixed) and input and output modality (ie, spoken, visual, and written). The eligible population included individuals who used chatbots to improve their psychological well-being or mental disorders but not physicians or caregivers who use chatbots for their patients. To be included, studies had to assess patients' perceptions and opinions about chatbots for mental health. The review included peer-reviewed articles, dissertations, and conference proceedings, and it excluded reviews, proposals, editorials, and conference abstracts. Only studies written in English were included in this review. No restrictions were applied regarding the study design, study setting, comparator, year of publication, or country of publication.

Textbox 1. Inclusion criteria.

- Intervention: chatbots operate as stand-alone software or a web browser
- Population: patients who use chatbots for improving their psychological well-being or mental disorders
- Outcome: patients' perceptions and opinions about mental health chatbots
- Type of publication: peer-reviewed articles, dissertations, and conference proceedings
- Language: English

Textbox 2. Exclusion criteria.

- Intervention: chatbots integrated into robotics, serious games, SMS, or telephone systems and those depend on human operator–generated dialog
- Population: physicians or caregivers who use chatbots for improving their psychological well-being or mental disorders
- Outcome: other outcomes
- Type of publication: reviews, proposals, editorials, and conference abstracts
- Language: other languages

Study Selection

In this review, MA and NA independently screened the titles and abstracts of all retrieved studies and independently read the full texts of studies included from the first step. AA resolved any disagreements between the reviewers. Cohen kappa was calculated to assess the intercoder agreement [22], which was 0.86 and 0.90 for screening titles and abstracts and for reading full texts, respectively, indicating excellent agreement [23].

Data Extraction

Multimedia Appendix 2 shows the data extraction form used in this review, which was pilot tested using 4 included studies. Data were extracted from the included studies by 2 reviewers independently (MA and NA), and the third reviewer (AA) resolved any discrepancies in decisions between the reviewers. Cohen kappa showed an excellent agreement (0.83) [23].

Assessment of Risk of Bias

Scoping reviews do not usually assess the risk of bias of the included studies because they have broad aims and include studies with diverse study designs [18,19]. The aim of this review was very broad, and the included studies had different study designs. Thus, the risk of bias of the included studies was not assessed in this review.

Data Synthesis

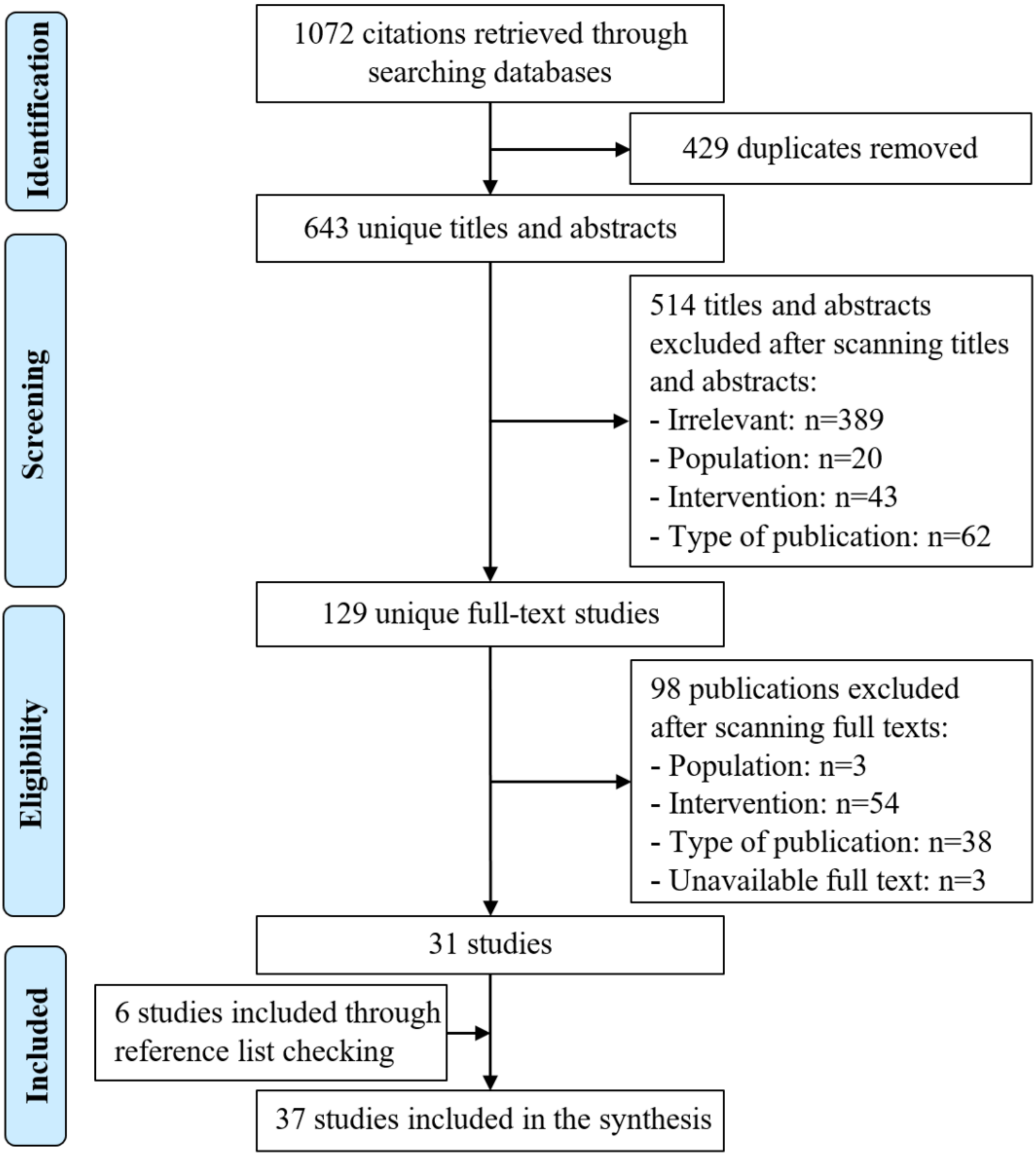
A narrative approach was used to synthesize the data extracted from the included studies. Thematic analysis was used to generate themes based on the findings of the included studies. This data synthesis approach (ie, thematic analysis) has been applied in numerous systematic and scoping reviews [24-28]. Given the exploratory nature of this review, an inductive approach was used to generate themes directly from the data [29]. The thematic analysis was conducted following the steps proposed by Braun and Clarke [30], as it is the most systematic guide for conducting thematic analysis to date [31,32]. Specifically, the first author (AA) scrutinized and rescrutinized the extracted data to familiarize himself with it. AA then coded the data systematically. On the basis of the assigned codes, themes were generated. All authors checked the fit of the generated themes to the extracted data and assigned codes. Where authors had differing opinions on the assigned codes and/or generated themes, these discrepancies were resolved through discussion. Finally, all authors participated in defining and naming the themes. Microsoft Excel was used to manage the analysis process.

Results

Search Results

As shown in Figure 1, 1072 citations were found by searching the electronic databases. After removing 429 duplicates of these citations, 643 titles and abstracts were screened. In the screening process, we excluded 514 titles and abstracts. Reading the full text of the remaining 129 citations resulted in a further 98 citations being excluded. The reasons for the exclusion are detailed in Figure 1. In backward and forward reference checking, we found 6 additional studies. In total, 37 studies were included in this review.

Figure 1. Flowchart of the study selection process.



Characteristics of Included Studies

As shown in Table 1, the most commonly used study design was a cross-sectional survey (34/37, 92%). About two-thirds of the studies were published as journal articles (24/37, 65%).

The included studies were conducted in more than 15 countries, but approximately 46% (17/37) of them were carried out in the United States. Approximately 62% (23/37) of the studies were published between 2015 and 2019.

Table 1. Characteristics of the included studies.

Parameters and characteristics	Studies[a]
Study metadata, n (%)	
Study design	
Survey	34 (92)
Quasi-experiment	2 (5)
Randomized controlled trial	1 (3)
Type of publication	
Journal article	24 (65)
Conference proceeding	12 (32)
Thesis	1 (3)
Country	
United States	17 (46)
Australia	3 (8)
France	3 (8)
The Netherlands	3 (8)
Japan	2 (5)
Germany	1 (3)
Korea	1 (3)
Spain	1 (3)
Sweden	1 (3)
Turkey	1 (3)
United Kingdom	1 (3)
Romania, Spain, and Scotland	1 (3)
Spain and Mexico	1 (3)
Global population	1 (3)
Year of publication	
Before 2010	3 (8)
2010-2014	11 (30)
2015-2019	23 (62)
Population characteristics	
Sample size, n (%)	
≤50	24 (65)
51-100	5 (14)
101-200	6 (16)
>200	2 (5)
Age (years)	
Mean (range)[b]	33.4 (13-79)
Sex , n (%)	
Male[c]	1436 (50)
Sample type, n (%)	
Clinical sample	21 (57)
Nonclinical sample	16 (43)

Parameters and characteristics	Studies[a]
Setting, n (%)[d]	
Clinical	14 (38)
Educational	12 (32)
Community	8 (22)
Intervention characteristics, n (%)	
Purpose	
Therapy	12 (32)
Training	9 (24)
Self-management	6 (16)
Counseling	5 (14)
Screening	4 (11)
Diagnosing	1 (3)
Platform	
Stand-alone software	24 (65)
Web based	13 (35)
Response generation	
Rule based	32 (86)
Artificial intelligence	5 (14)
Dialog initiative	
Chatbot	32 (86)
Both	5 (14)
Embodiment	
Yes	30 (81)
No	7 (19)
Targeted disorders[e]	
Depression	41 (23)
Autism	6 (16)
Anxiety	6 (16)
Any mental disorder	6 (16)
Substance use disorder	5 (14)
Posttraumatic stress disorder	5 (14)
Schizophrenia	3 (8)
Stress	3 (8)

[a] Percentages were rounded and may not sum to 100.

[b] Mean age was reported in 24 studies.

[c] Sex was reported in 29 studies.

[d] Setting was reported in 34 studies.

[e] Numbers do not add up as several chatbots target more than one health condition.

The sample size was 50 or less in 24 studies and more than 200 in 2 studies (Table 1). The participants' age was reported in 24 studies, with a mean age of participants was 33.4 years (SD 15.2; range 13-79 years). The sex of participants was reported in 29 studies, where the mean percentage of men was 49.5%. In 57% (21/37) of the studies, participants were from clinical populations (ie, they had mental disorders). The sample was recruited from clinical (n=14), educational (n=12), or community settings (n=8). Multimedia Appendix 3 shows the metadata and population characteristics of each included study.

The 37 included studies assessed patients' perceptions and opinions about 32 different chatbots. Chatbots were used for therapeutic purposes (n=12), training (n=9), self-management (n=6) counseling (n=5), screening (n=4), and diagnosis (n=1; Table 1). Chatbots were implemented in stand-alone software in 65% (24/37) of the studies, whereas the remaining chatbots were implemented in web-based platforms. In the majority of studies (32/37, 86.5%), chatbots generated their responses based on predefined rules or decision trees (rule based). Chatbots in the remaining studies generated their responses based on machine learning approaches. Chatbots led the dialog in most studies (n=32/37, 86.5%); both chatbots and users could lead the dialog in the remaining studies. Virtual agents (eg, avatar or virtual human) were embodied in chatbots in 30 of 37 studies (81.1%). The most common disorder targeted by chatbots was depression (n=15, 40.5%). Multimedia Appendix 4 shows the characteristics of the intervention in each included study [33-69].

Study Findings

The thematic analysis generated 10 themes from the findings of the studies: usefulness, ease of use, responsiveness, understandability, acceptability, attractiveness, trustworthiness, enjoyability, content, and comparisons. More details about these themes are elaborated in the following subsections.

Theme 1: Usefulness

In total, 20 studies investigated the usefulness of chatbots and/or their features for patients [33-52]. In 3 studies [41,47-49,51], the overall usefulness of chatbots was rated as high. Participants reported that chatbots are useful for practicing conversations in a private place [33,46], learning [37,38,42,46], making users feel better [40], preparing users for interactions with health care providers [43], implementing the learned skills in daily life [46], facilitating a sense of accountability from daily check-in [37,38], keeping the learned skills more prominently in users' minds [46], increasing users' self-efficacy [46], improving users' self-confidence and readiness for job interviews [47-49], and recalling users' memories [51]. However, participants in one study doubted the usefulness of chatbots for smoking cessation [39].

Users considered the following components of chatbots useful: real-time feedback [33,45,50], diary [52], weekly summary [42], presenting the helpline during the conversation [36], and psychoeducation [52]. Some studies found that virtual agents embodied in chatbots are useful for motivating users to use chatbots [34], establishing a relationship with them [35], and feeling supported by them [45]. However, other studies demonstrated that participants had neutral perceptions and opinions about the added value of embodied virtual agents with speech [52] or without speech [44].

Theme 2: Ease of Use

The ease of use and usability of chatbots were assessed in 20 studies [33,34,36,39,43,46-51,53-61]. Participants in 15 studies rated the overall ease of use of chatbots as high [36,39,43,47-50,53-55,57-61]. A total of 5 studies assessed the usability of chatbots [34,36,46,51,56], and it was rated high in all these studies (ranging from 69 to 88.2). Participants in 3 studies reported that chatbots are easy to learn and become familiar with [33,39,55]. Although participants did not find chatbots difficult to navigate in one study [33], they rated the navigation of the chatbot as low in another study [36].

In 3 studies, participants faced difficulty in using the chatbot because they did not know when [60] and how [61] to reply to it, and they did not have enough options to reply to it [62]. Some participants in 3 studies criticized chatbots because of technical glitches that happened during their use, such as screen freezing [37,45,60]. Users suggested that chatbots allow them to respond by speaking and not typing to make them easy to use [57].

Theme 3: Responsiveness

This theme brings together perceptions and opinions of participants about verbal and nonverbal responses generated by chatbots in terms of realism, repetitiveness (variability), speed, friendliness, and empathy. A total of 10 studies assessed participants' perceptions and opinions about how real the chatbots were in terms of verbal and nonverbal responses. Although participants in 7 studies had mixed or neutral perceptions and opinions about the realism of verbal and nonverbal responses [33,35,38,45,46,50,56], most participants in 3 studies held corresponding positive perceptions and opinions [52,57,60]. Participants believed that chatbots may be more realistic if they have an embodied virtual agent [44] and are able to communicate verbally [33].

Most participants in several studies stated that chatbots were able to show friendly [34,52,59,60,63] and emotional [35,37,38] responsiveness. However, participants in other studies had mixed perceptions and opinions about the ability of chatbots to generate friendly [35,44,64] and emotional [64] responses. Participants in one study found chatbots with an embodied virtual agent friendlier than those without an embodied virtual agent [44].

A total of 7 studies demonstrated that chatbot responses were repetitive [35-38,40,42,61]. Participants in one study reported that the repetitiveness of responses made the chatbot look less real [42]. Moreover, 3 studies concluded that the speed of chatbot responses was appropriate [57,60,61]. Yet, participants in 2 studies criticized the speed of chatbot responses as either too fast [38] or too slow [33]. In 6 studies, participants did not face any difficulties in understanding and interpreting chatbot responses [45,57,59,60,65,66].

In general, participants in 6 studies were satisfied with chatbot responses [33-36,62,63,67]. Participants attributed their satisfaction to the accuracy and consistency of chatbot feedback with what their therapist recommended in the past [33], appropriate use of high-quality elements (eg, Graphics Interchange Format images) [36], consistency of voice and tone of the chatbot [36], and quality of information provided [62,63]. However, participants in 4 studies were not satisfied by the conversation of chatbot because of confusing responses [57], disturbing users [40], the robotic quality of its voice [62], using emoticons (emojis) [37], conversations being too short [37], and shallowness of the conversations [42].

Participants suggested several enhancements related to the responsiveness of chatbots, such as the ability to speak [44], more flowing speech [33], friendlier voice [44], immediate responses [33,45], faster blinks [33], more explanation [33], providing feedback on each answered question [44], and more precise feedback [45].

Theme 4: Understandability

This theme brings together perceptions and opinions of participants about the ability of chatbots to understand their verbal and nonverbal contact. Chatbot understandability for verbal responses was rated as high among participants in 3 studies [33,45,61]; however, participants in other studies criticized the inability of the chatbot to understand their verbal responses in general [36-38,40,57], misspelled words (eg, anious instead of anxious) [36], different languages [36], unexpected answers [37,38], and nonverbal responses (eg, images, emojis, and facial expressions) [33,36,56]. Therefore, participants in 3 studies suggested that future chatbots should better understand and recognize their verbal and nonverbal responses [33,37,43].

Theme 5: Acceptability

This theme concerns participants' acceptability of chatbots and its functionalities and their intentions to use them in the future. The acceptability of chatbots was rated high by users in 12 studies [34,37,38,43,45,46,53,54,57,61,66,68]. Wearing eye tracker goggles [62] or the inclusion of an embodied virtual agent [65] was acceptable for participants. There were mixed opinions about the acceptability of chatbots in one study [35]. Chatbots connected with a perception system (camera) for extracting data about user behavior was not acceptable for most participants in one study [60]. Users in one study indicated that the acceptability of chatbots could increase if the purpose of chatbots was clearly explained [33]. Note that the lack of clarity of the purpose of chatbots was highlighted by participants in 2 studies [33,42].

Furthermore, 6 studies demonstrated that people would like to use chatbots in the future [35,43,45,46,50,53,54,61,65], whereas participants in 2 studies were ambivalent about the future use of chatbots [33,39]. This ambivalence resulted from participants perceiving chatbots as a conversational tool for children [33]. Some participants reported that they probably would not use chatbots unless they popped up on their screens [33]. Although participants in one study showed high intention to use the chatbot in the future, they highlighted that it is inconvenient or inappropriate to use it every day [35].

Theme 6: Attractiveness

Participants in one study rated the attractiveness of a chatbot as low [57]. Furthermore, Demirci [55] found that perceptions and opinions of users about attractiveness changed from positive before using the chatbot to neutral after using it. Participants suggested improvements of the controls (eg, icons, combo boxes, buttons, and font size) [33,34] and the appearance of the embodied virtual agent [34] to obtain a more attractive graphical user interface.

Theme 7: Trustworthiness

This theme concerns participants' trust in chatbot. In 7 studies, participants believed that chatbots are trustworthy [34,44,46,52,54,57,63]. One study concluded that participants were satisfied with the anonymity, confidentiality, and objectivity of chatbots [63]. Most participants in the 2 studies were comfortable to share and did share personal information with the chatbot [56,63].

Theme 8: Enjoyability

Participants in 9 studies considered using chatbots as enjoyable and fun [36,42-44,47-49,52,55,59]. However, one study found that perceptions and opinions of users about enjoyment changed from positive before using the chatbot to neutral after using it [55].

Theme 9: Content

This theme contains participants' opinions about the content of chatbots. In 6 studies, participants were satisfied with the contents of chatbots such as videos, games, topics, suggestions, and weekly graphs [34,37,38,42,43,52]. However, the content of chatbots was criticized by users because of its superficiality [33,38], irrelevancy [38], having too long videos [37], and having overwhelming information [57]. Participants in 3 studies appreciated the personalization feature in chatbots that allowed them to customize the session length and the gender and appearance of the virtual agent [35,57,60]. Participants suggested that chatbots should contain psychoeducation [35], more therapy-based activities [34], longer and more frequent intervention sessions [43], more customizable content [35,43], and interventions for other chronic health conditions [43].

Theme 10: Comparisons

This theme brings together participant perspectives about chatbots in comparison with other chatbots or traditional methods. Although most participants in one study preferred interacting with a chatbot rather than a human for their health care [53], participants in another 2 studies had mixed preferences about that [33,45]. In 2 studies [44,58], participants preferred using chatbots with an embodied virtual agent compared with chatbots without an embodied virtual agent.

Participants in one study preferred that chatbot provides real-time feedback on their nonverbal behavior rather than postsession feedback [33]. According to Morris et al [67], participants preferred the chatbot's responses drawn from an existing pool of web-based peer support data rather than those generated by the chatbot itself. Users preferred to use chatbots that provide information in an interactive fashion [43], are added on a device that they already own (eg, smartphones, tablets, computers) [60], and call them by their first name [60].

A chatbot without an embodied virtual agent (text-based chatbot) was compared with 2 chatbots with an embodied virtual agent (one reacts to the user with verbal and nonverbal empathic reactions, whereas the other did not) in another study [58]. The study found that the empathic chatbot was more acceptable, enjoyable, empathizing, understanding, nicer, sociable, trustworthy, realistic, private, anthropomorphic, animated, intelligent, socially influencing, friendlier, and safer than the nonempathic chatbot and the text-based chatbot [58]. Furthermore, the study demonstrated that participants are more willing to disclose information to the text-based chatbot than to empathic and nonempathic chatbots and a human counselor [58]. The study also found that participants were more willing to use empathic chatbots than nonempathic chatbots and text-based chatbots [58].

One study compared AI chatbots with an individual or a chatbot controlled by the same individual (Wizard-of-Oz) [56]. The study found that the Wizard-of-Oz chatbot was rated by participants as more usable and listened better than the AI chatbot [56]. Furthermore, users of the Wizard-of-Oz chatbot felt greater rapport than users of the AI chatbot and, surprisingly, than those who were interviewed by humans [56]. However, there was no difference between users of the AI chatbot and those interviewed by a human in their ratings of rapport [56].

In another study [69], participants felt a greater rapport with the real expert than with a rule-based chatbot. Participants also indicated that the rule-based chatbot is less able to understand their responses and feelings than a real expert [69]. Furthermore, participants found the real expert more trustworthy than the rule-based chatbot [69]. Participants reported that the chatbot is more enjoyable and engaging than traditional methods of learning and practicing dialectical behavior therapy skills [46].

Discussion

Principal Findings

The main finding of this review is that there are features of chatbots that health care providers cannot deliver over a long period. These features have been identified as useful in mental health chatbots: real-time feedback, weekly summary, and continuous data collection in terms of a diary. Usefulness and ease of use are aspects of chatbots that have been studied most comprehensively in the analyzed papers. Overall, the usefulness of mental health chatbots is perceived as high by patients. According to these studies, patients find chatbot systems easy to use. Interactional enjoyment and perceived trust are significant mediators of chatbot interaction [70]. Although chatbots are perceived as useful and easy to use, participants of reported studies recognized the existing conversational limitations of those systems: conversations are perceived as shallow, confusing, or too short. This points to an important issue to be addressed in future mental health chatbot development (see the *Practical and Research Implications* section). The conversation quality still needs to be improved. In this context, chatbot quality in terms of responsiveness and variability of responses is an important issue. Currently, systems are rather restricted in the number of responses, which might be because of the early development stage of many chatbots, as has already been reported by Laranjo et al [71]. Another relevant aspect judged important is the quality of provided information and consistency with recommendations of treating physicians. The implications of these principal findings toward practice and research are described in the *Practical and Research Implications* section.

Comparison With Existing Literature

This is the first review that summarizes perceptions and opinions of patients about mental health chatbots, as reported by previous studies. Palanica et al [72] assessed physicians' perceptions of health chatbots. They found that physicians see the benefits of those apps basically in support of patients in managing their health and improving physical, psychological, and behavioral health outcomes. As limitations of health chatbots, they mentioned the inability to comprehend and express human emotions and a lack of medical expert knowledge and intelligence. Our results provide the counterpart of this observation, namely, patients recognizing limitations in the conversation quality of health chatbots. A recent paper on a chatbot for regulating emotions shows that perceptions of patients and psychologists differ regarding attractiveness, efficiency, and stimulations of health chatbots [73]. Although psychologists see these aspects rather positive, patients are more critical and complain about the restrictions of chatbot conversations.

In their review of the landscape of psychiatric chatbots, Vaidyam [4] identified studies that showed high satisfaction with the interventions they received. Participants reported the interventions as helpful, easy to use, and informative and rated satisfaction highly (>4.2 out of 5) on all scales, including ease of use, desire to continue using the system, liking, and trust. In addition, the results of Bendig et al [16] suggest the practicability, feasibility, and acceptance of using chatbots to promote mental health. Our results confirm these observations: patients consider chatbots as useful, and acceptability is rated high.

A study assessed the use of mobile technologies in health-related areas from various perspectives [74]. It points to the following risks for mHealth app usage: lack of functionality, dissemination of false information, misdiagnosis, mistreatment, and unknown unwanted side effects. From the patient perspective, these issues could also be identified in our review: quality of provided information and consistency with recommendations of treating physicians are relevant aspects. The results of the study by Albrecht [74] go beyond our results by pointing to the risks of mHealth apps in case of an emergency. Implemented algorithms still lack reactivity (eg, in the case of self-endangerment or hazards of others). In addition, Singh et al [75] showed that only 23% of mHealth apps responded adequately to dangerous user input (eg, suicidal ideations). This illustrates the enormous need for improvement in terms of responsiveness of mHealth apps in potentially dangerous situations. According to our results, the patients did not seem to be too concerned about this issue of chatbot behavior in emergencies. It was not mentioned in the reviewed studies.

Practical and Research Implications

Practical Implications

The study results have the following practical implications. To be useful, we need to create high-quality chatbots that are able to respond to a user in multiple ways. A mental health chatbot must be empathic to be perceived as motivating and engaging and to establish a relationship with the user. A study by de Gennaro [76] supports this by demonstrating that empathic chatbots have the potential to provide emotional support to victims of social exclusion.

The patient-doctor or patient-therapist relationship in standard health care settings is characterized by trust and loyalty. Measurements must be undertaken to make the chatbot-patient relationship also trustworthy. This could be realized by providing information on the secondary use of the collected patient data on data storage and analysis procedures. Another approach is blended therapy [77], a combination of face-to-face and web-based or digital therapy, which shows the potential of a cost-effective and accessible format in cognitive behavioral therapy. This would also address another practical implication, which is that the chatbot has to be related to the therapy. In particular, recommendations provided by a chatbot must be consistent with the recommendations of the treating health care professionals. This demands the integration of chatbots into the health care process; the chatbot should be aware of the recommendations or treatment plans of the health care provider. Finally, to increase the acceptance of chatbot use in patients, physicians need to be convinced of the usefulness of those systems so that they will recommend them to patients. Studies suggest that there are already physicians who are convinced of the usefulness [72]. Given the strong bond of trust among patients toward their physicians, patients will be convinced of the usefulness of an app once their physician recommends it.

Research Implications

From the practical implications, we can derive the following research implications. There is still a need to improve the linguistic capabilities of mental health chatbots [71]. Their ability to understand and react appropriately to user input has to be increased. Furthermore, methods are required to generate dynamic answers to ensure the variability of chatbot responses. Linguistic or lexical variability can be added to the knowledge base of rule-based chatbots, but the capabilities will always depend on the completeness of the knowledge base. Methods for slightly adapting or reformulating responses from the knowledge base could help in addressing this issue. In domains outside the health care domain, crowdsourcing has been applied to improve conversation quality [78]. However, in health care, we have to be careful with learning from data because we have to ensure that responses and recommendations are in line with clinical evidence. It is still an open research question on how to learn clinical evidence to train health chatbots.

Furthermore, methods have to be developed to deal with unexpected user input and to detect critical situations. In mental health, it is crucial to react appropriately for people who are at risk of suicide or self-harm [79]. Sentiment analysis methods have proven to be successful in analyzing social media messages with respect to suicide and self-harm [80]. These methods could be useful in health chatbots as well. The main challenge is the use of an appropriate reaction once an emergency situation has been detected. Another interesting research topic is the customization or personalization of chatbots to individual users. This topic is still in its infancy [81]. As long as mental health chatbots rely on decision trees or fix implemented rule bases, they will not be able to adapt to specific user needs. We can construct the knowledge base in such a way that there are responses for different types of users, but this will be time consuming and will always be incomplete. Learning from conversations with users could help. The style or complexity of language can be adapted based on the given user input. Patient-specific knowledge, for example, on treatment plans could be retrieved from health records. Methods are required to include such knowledge dynamically to a chatbot. In this way, the chatbot content is adapted to match individual needs.

For evaluating the mental health chatbot, benchmarks have to be created, and consistent metrics and methods have to be developed. Laranjo et al [71] reviewed the characteristics, current applications, and evaluation measures of health chatbots. Evaluation measures were divided into 3 main types: technical performance, user experience, and health research measures. The first attempts toward evaluation frameworks for digital health interventions [82] and health chatbots [83,84] have been recently published. Depending on the facet under consideration, different metrics can be used: system performance and effectiveness can be evaluated using different computational metrics (eg, usability, ease of use, usefulness) [85]. Software quality can be measured by reliability, security, maintainability, and efficiency using software engineering metrics [86]. If the system uses AI and machine learning techniques, the metrics comprise the accuracy and precision of predictions and recommendations. Furthermore, the efficiency of the systems has to be evaluated and compared with existing care models. With regard to safe app use, 3 criteria should be evaluated: (1) quality of the therapeutic content, (2) functionality, and (3) data safety and protection [87].

Strengths and Limitations

Strengths

This review was developed, executed, and reported according to the PRISMA Extension for Scoping Reviews [20]. This enabled us to produce a high-quality review.

The most commonly used databases in health and information technology were searched to retrieved relevant studies as many as possible. Searching Google Scholar and carrying out backward and forward reference list checking enabled us to identify gray literature and minimize the risk of publication bias as much as possible. As no restrictions were applied regarding the study design, study setting, comparator, year of publication, and country of publication, this review can be considered comprehensive.

Selection bias in this review was minimal because study selection and data extraction were performed independently by 2 reviewers. Furthermore, the agreement between reviewers was very good for study selection and data extraction. This study is one of the few reviews that used thematic analysis to synthesize the findings of the included studies. The thematic analysis followed the highly recommended guidelines proposed by Braun and Clarke [30].

Limitations

This review focused on chatbots that only work on stand-alone software and a web browser (but not robotics, serious games, SMS, or telephones). Furthermore, this review was restricted to chatbots that are not controlled by human operators (Wizard-of-Oz). Therefore, perceptions and opinions of patients found in this review may be different from their perceptions and opinions about Wizard-of-Oz chatbots and/or chatbots with alternative modes of delivery. The abovementioned restrictions were applied by previous reviews about chatbots, as these features are not part of ordinary chatbots [4,14,17].

Owing to practical constraints, we restricted the search to English studies and we could not search interdisciplinary databases (eg, Web of Science and ProQuest), conduct manual search, or contact experts. Consequently, it is likely that we have missed some English and non-English studies. Most included studies were conducted in developed countries, particularly in the United States. Therefore, the findings of this review may not be generalizable to developing countries, as patients in such countries may have different perceptions and opinions about mental health chatbots.

Conclusions

In this paper, we explored perceptions and opinions of patients about mental health chatbots, as reported in the existing literature. The results demonstrated that there are overall positive perceptions and opinions of patients about mental health chatbots, although there is some skepticism toward trustworthiness and usefulness. Many important aspects have been identified to be addressed in research and practice. Among them are the need to improve the linguistic capabilities of chatbots and seamless integration into the health care process. Future research will have to pick up those issues to create successful, well-perceived chatbot systems, and we will start developing corresponding concepts and methods. The research implications are also relevant for health care chatbots beyond mental health chatbots. Their consideration has the potential to improve patients' perceptions of health care chatbots in general.

Acknowledgments

The publication of this study was funded by the Qatar National Library. This study was a part of a project funded by the Qatar National Research Fund (NPRP12S-0303-190204). The project title is *A Personalized and Intelligent Digital Mental Health Platform for Qatar and the Arab world*.

Authors' Contributions

AA developed the protocol and conducted a search with guidance from and under the supervision of MH and BB. Study selection and data extraction were performed independently by MA and NA. AA executed the analysis, and all authors checked the validity of the generated themes. AA and KD drafted the manuscript, and it was revised critically for important intellectual content by all authors. All authors approved the manuscript for publication and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 28 KB - jmir_v23i1e17828_app1.docx]

Multimedia Appendix 2
Data extraction form.
[DOCX File , 19 KB - jmir_v23i1e17828_app2.docx]

Multimedia Appendix 3
The metadata and population characteristics of each included study.
[DOCX File , 30 KB - jmir_v23i1e17828_app3.docx]

Multimedia Appendix 4
Characteristics of the intervention in each included study.
[DOCX File , 29 KB - jmir_v23i1e17828_app4.docx]

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
2. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. PLoS One 2015;10(2):e0116820 [FREE Full text] [doi: 10.1371/journal.pone.0116820] [Medline: 25658103]

3. Jones SP, Patel V, Saxena S, Radcliffe N, Ali Al-Marri S, Darzi A. How Google's 'Ten Things We Know To Be True' could guide the development of mental health mobile apps. Health Aff (Millwood) 2014 Sep;33(9):1603-1611. [doi: 10.1377/hlthaff.2014.0380] [Medline: 25201665]
4. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
5. Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. BJPsych Int 2016 Aug;13(3):61-63 [FREE Full text] [doi: 10.1192/s2056474000001240] [Medline: 29093905]
6. Murray C, Vos T, Lozano R, Naghavi M, Flaxman A, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
7. Hester RD. Lack of access to mental health services contributing to the high suicide rates among veterans. Int J Ment Health Syst 2017;11:47 [FREE Full text] [doi: 10.1186/s13033-017-0154-2] [Medline: 28828036]
8. Blankers M, Schippers G. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. Suchttherapie 2011 Aug 22;12(S 01). [doi: 10.1055/s-0031-1284519]
9. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
10. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Front. Robot. AI 2017 Oct 12;4. [doi: 10.3389/frobt.2017.00051]
11. Sebastian J, Richards D. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. Comput Hum Behav 2017 Aug 2;73(3):479-488. [doi: 10.1016/j.chb.2017.03.071]
12. Hawton K, van Heeringen K. Suicide. Lancet 2009 Apr;373(9672):1372-1381 [FREE Full text] [doi: 10.1016/s0140-6736(09)60372-x]
13. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 7;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
14. Abd-Alrazaq AA, Bewick B, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
15. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick B, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
16. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Verhaltenstherapie 2019 Aug 20:1-13. [doi: 10.1159/000501812]
17. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 9;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Method 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
19. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 4;169(7):467. [doi: 10.7326/m18-0850]
21. Search Tip. Google Scholars. 2020. URL: https://scholar.google.com/intl/en/scholar/help.html [accessed 2020-12-21]
22. Higgins J, Deeks J. Selecting Studies and Collecting Data. London, UK: John Wiley & Sons; 2008.
23. Orwin R, Vevea J. Evaluating coding decisions. In: The Handbook of Research Synthesis and Meta-analysis. Second Edition. New York, USA: Russell Sage Foundation; 2009:177-203.
24. Caton S, Chapman M. The use of social media and people with intellectual disability: a systematic review and thematic analysis. J Intellect Dev Disable 2016 Mar 9;41(2):125-139. [doi: 10.3109/13668250.2016.1153052]
25. Crampton NH, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. J Am Med Inform Assoc 2016 May;23(3):654-665. [doi: 10.1093/jamia/ocv178] [Medline: 26769911]
26. Darbyshire D, Baker P. A systematic review and thematic analysis of cinema in medical education. Med Humanit 2012 Jun;38(1):28-33. [doi: 10.1136/medhum-2011-010026] [Medline: 22282424]
27. Freeman J, Caldwell P, Bennett P, Scott K. How adolescents search for and appraise online health information: a systematic review. J Pediatr 2018 Apr;195:244-55.e1. [doi: 10.1016/j.jpeds.2017.11.031] [Medline: 29398062]
28. Smith V, Begley CM, Clarke M, Devane D. Professionals' views of fetal monitoring during labour: a systematic review and thematic analysis. BMC Pregnancy Childbirth 2012 Dec 27;12:166 [FREE Full text] [doi: 10.1186/1471-2393-12-166] [Medline: 23270400]
29. Boyatzis R. Transforming qualitative information: thematic analysis and code development. London, UK: Sage Publications; 1998:302.

30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
31. Howitt D. Introduction to Qualitative Research Methods in Psychology. London, UK: Pearson; 2016.
32. Wilson S, MacLean R. Research Methods and Data Analysis for Psychology. New York, USA: McGraw-Hill Higher Education; 2011.
33. Ali M, Rasazi Z, Mamun A, Langevin R, Rawassizadeh R, Schubert L. A virtual conversational agent for teens with autism: experimental results and design lessons. arXiv 2018:- epub ahead of print. [doi: 10.1145/3383652.3423900]
34. Bresó A, Martínez-Miranda J, Botella C, Baños RM, García-Gómez JM. Usability and acceptability assessment of an empathic virtual agent to prevent major depression. Expert Syst 2016 May 25;33(4):297-312. [doi: 10.1111/exsy.12151] [Medline: 2016]
35. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
36. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S. Assessing the Usability of a Chatbot for Mental Health Care. In: International Conference on Internet Science. 2019 Presented at: CIS'19; October 24-26, 2018; St. Petersburg, Russia p. 121-132. [doi: 10.1007/978-3-030-17705-8_11]
37. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
38. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
39. Grolleman J, van DB, Nijholt A, van EA. Break the Habit! Designing an e-Therapy Intervention Using a Virtual Coach in Aid of Smoking Cessation. In: International Conference on Persuasive Technology. 2006 Presented at: CPT'06; May 18-19, 2006; Eindhoven, The Netherlands p. 133-141. [doi: 10.1007/11755494_19]
40. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
41. Ku J, Han K, Lee HR, Jang HJ, Kim KU, Park SH, et al. VR-based conversation training program for patients with schizophrenia: a preliminary clinical trial. Cyberpsychol Behav 2007 Aug;10(4):567-574. [doi: 10.1089/cpb.2007.9989] [Medline: 17711366]
42. Ly K, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
43. Pinto MD, Greenblatt AM, Hickman T, Rice HM, Thomas TL, Clochesy JM. Assessing the critical parameters of esmart-MH: a promising avatar-based digital therapeutic intervention to reduce depressive symptoms. Perspect Psychiatr Care 2016 Jul;52(3):157-168. [doi: 10.1111/ppc.12112] [Medline: 25800698]
44. Pontier M, Siddiqui G. A Virtual Therapist That Responds Empathically to Your Answers. In: International Workshop on Intelligent Virtual Agents. 2008 Presented at: WIVA'08; September 1-3, 2008; Tokyo, Japan p. 214-225. [doi: 10.1007/978-3-540-85483-8]
45. Razavi S, Ali M, Smith T, Schubert L, Hoque M. The LISSA Virtual Human and ASD Teens: An Overview of Initial Experiment. In: International Conference on Intelligent Virtual Agents. The LISSA virtual human and ASD teens; 2016 Presented at: CIVA'16; September 20-23, 2016; Los Angeles, USA. [doi: 10.1007/978-3-319-47665-0_55]
46. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 12-14, 2018; Montreal, Canada. [doi: 10.1145/3173574.3173972]
47. Poole-Wilson PA, Langer GA. Effect of pH on ionic exchange and function in rat and rabbit myocardium. Am J Physiol 1975 Sep;229(3):570-581. [Medline: 2014]
48. Smith MJ, Ginger EJ, Wright K, Wright MA, Taylor JL, Humm LB, et al. Virtual reality job interview training in adults with autism spectrum disorder. J Autism Dev Disord 2014 Oct;44(10):2450-2463 [FREE Full text] [doi: 10.1007/s10803-014-2113-y] [Medline: 24803366]
49. Smith MJ, Humm LB, Fleming MF, Jordan N, Wright MA, Ginger EJ, et al. Virtual reality job interview training for veterans with posttraumatic stress disorder. J Vocat Rehabil 2015;42(3):271-279 [FREE Full text] [doi: 10.3233/JVR-150748] [Medline: 27721645]
50. Tanaka H, Sakti S, Neubig G, Toda T, Negoro H, Iwasaka H. Automated Social Skills Trainer. In: Proceedings of the 20th International Conference on Intelligent User Interfaces. 2015 Presented at: CIUI'15; 29 March-01 April 2015; Atlanta Georgia, USA. [doi: 10.1145/2678025.2701368]
51. Tielman ML, Neerincx MA, Bidarra R, Kybartas B, Brinkman W. A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. J Med Syst 2017 Aug;41(8):125 [FREE Full text] [doi: 10.1007/s10916-017-0771-y] [Medline: 28699083]

52. Tielman ML, Neerincx MA, van Meggelen M, Franken I, Brinkman W. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. Technol Health Care 2017 Dec 4;25(6):1081-1096 [FREE Full text] [doi: 10.3233/THC-170899] [Medline: 28800346]
53. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
54. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
55. Demirci H. User experience over time with conversational agents case study of woebot on supporting subjective well-being. Ankara, Turkey: Middle East Technical University; 2018.
56. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A. Simsensei Kiosk: a Virtual Human Interviewer for Healthcare Decision Support. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. 2014 Presented at: AAMS'14; 3-5 May 2014; Paris, France.
57. Elmasri D, Maeder A. A Conversational Agent for an Online Mental Health Intervention. In: International Conference on Brain Informatics. 2016 Presented at: CBI'16; 13-16 October 2016; Omaha, USA. [doi: 10.1007/978-3-319-47103-7_24]
58. Lisetti C, Amini R, Yasavur U, Rishe N. I can help you change! An empathic virtual agent delivers behavior change health interventions. ACM Trans Manage Inf Syst 2013 Dec 1;4(4):1-28. [doi: 10.1145/2544103]
59. Milne M, Luerssen M, Lewis T, Leibbrandt R, Powers D. Development of a Virtual Agent Based Social Tutor for Children With Autism Spectrum Disorders. The 2010 International Joint Conference on Neural Networks (IJCNN). In: The 2010 International Joint Conference on Neural Networks. 2010 Presented at: IJCNN'10; July 2-7, 2010; Barcelona, Spain p. 18-23. [doi: 10.1109/IJCNN.2010.5596584]
60. Wargnier P, Benveniste S, Jouvelot P, Rigaud A. Usability assessment of interaction management support in LOUISE, an ECA-based user interface for elders with cognitive impairment. TAD 2018 Nov 26;30(3):105-126 [FREE Full text] [doi: 10.3233/tad-180189]
61. Yasavur U, Lisetti C, Rishe N. Let’s talk! speaking virtual counselor offers you a brief intervention. J Multimodal User Interfaces 2014 Sep 5;8(4):381-398. [doi: 10.1007/s12193-014-0169-9]
62. Luerssen M, Hawke T. Virtual Agents as a Service: Applications in Healthcare. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. 2018 Presented at: CIVA'18; 20-23 November 2018; Sydney, Australia. [doi: 10.1145/3267851.3267858]
63. Swartout W, Artstein R, Forbell E, Foutz S, Lane HC, Lange B, et al. Virtual Humans for Learning. AIMag 2013 Dec 15;34(4):13-30. [doi: 10.1609/aimag.v34i4.2487]
64. Martínez-Miranda J, Bresó A, García-Gómez J. Look on the bright side: a model of cognitive change in virtual agents. In: International Conference on Intelligent Virtual Agents. 2014 Presented at: CIVA'14; 26-29 August 2014; Boston, MA, USA p. 285-294. [doi: 10.1007/978-3-319-09767-1_37]
65. Lahiri U, Bekele E, Dohrmann E, Warren Z, Sarkar N. Design of a virtual reality based adaptive response technology for children with autism. IEEE Trans Neural Syst Rehabil Eng 2013 Jan;21(1):55-64 [FREE Full text] [doi: 10.1109/TNSRE.2012.2218618] [Medline: 23033333]
66. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7:42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
67. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
68. Auriacombe M, Moriceau S, Serre F, Denis C, Micoulaud-Franchi JA, de Sevin E, et al. Development and validation of a virtual agent to screen tobacco and alcohol use disorders. Drug Alcohol Depend 2018 Dec 1;193:1-6. [doi: 10.1016/j.drugalcdep.2018.08.025] [Medline: 30321739]
69. Yokotani K, Takagi G, Wakashima K. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. Comput Hum Behav 2018 Aug;85(6):135-145. [doi: 10.1016/j.chb.2018.03.045] [Medline: 2018]
70. Fadhil A, Schiavo G. Designing for Health Chatbots. ArXiv 2019:1-10.
71. Laranjo L, Dunn A, Tong H, Kocaballi A, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
72. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 5;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
73. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. IEEE Trans Emerg Topics Comput 2020:1-1. [doi: 10.1109/tetc.2020.2974478]
74. Albrecht U. Chances and Risks of Mobile Health Apps (CHARISMHA). 2016 May 26. URL: https://nbn-resolving.org/urn:nbn:de:gbv:084-16051809293 [accessed 2020-12-21]

75. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 1;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
76. de Gennaro M, Krumhuber EG, Lucas G. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. Front Psychol 2019;10:3061 [FREE Full text] [doi: 10.3389/fpsyg.2019.03061] [Medline: 32038415]
77. Fitzpatrick M, Nedeljkovic M, Abbott J, Kyrios M, Moulding R. 'Blended' therapy: the development and pilot evaluation of an internet-facilitated cognitive behavioral intervention to supplement face-to-face therapy for hoarding disorder. Internet Interv 2018 Jun;12:16-25 [FREE Full text] [doi: 10.1016/j.invent.2018.02.006] [Medline: 30135765]
78. Yu Z, Xu Z, Black AW, Rudnicky A. Chatbot Evaluation and Database Expansion via Crowdsourcing. In: Proceedings of the chatbot workshop of LREC. 2016 Presented at: LREC'16; 2016; UK.
79. Luxton DD. Ethical implications of conversational agents in global public health. Bull World Health Organ 2020 Jan 27;98(4):285-287. [doi: 10.2471/blt.19.237636]
80. Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, Wiebe J, et al. Sentiment analysis of suicide notes: a shared task. Biomed Inform Insights 2012 Jan 30;5(Suppl 1):3-16 [FREE Full text] [doi: 10.4137/bii.s9042] [Medline: 22419877]
81. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 7;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
82. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. Inf Technol 2019;61(5-6):253-263. [doi: 10.1515/itit-2019-0019]
83. Denecke K, Warren J. How to evaluate health applications with conversational user interface? Stud Health Technol Inform 2020 Jun 16;270:976-980. [doi: 10.3233/SHTI200307] [Medline: 32570527]
84. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. J Med Internet Res 2020 Jun 5;22(6):e18301 [FREE Full text] [doi: 10.2196/18301] [Medline: 32442157]
85. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
86. Kan S. Metrics and Models in Software Quality Engineering. Boston: USA: Addison-Wesley Longman Publishing Co., Inc; 2002.
87. Neary M, Schueller S. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]

Abbreviations

AI: artificial intelligence
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by M Leary, M Lahti, YC Wang, E Broglia, J Apolinário-Hagen; comments to author 24.03.20; revised version received 01.06.20; accepted 21.06.20; published 13.01.21.

Please cite as:
Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M
Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review
J Med Internet Res 2021;23(1):e17828
URL: http://www.jmir.org/2021/1/e17828/
doi:10.2196/17828
PMID:33439133

Original Paper

Deliberative and Paternalistic Interaction Styles for Conversational Agents in Digital Health: Procedure and Validation Through a Web-Based Experiment

Theresa Schachner[1*], BSc, BA, MSc; Christoph Gross[1*], BSc, MA; Andrea Hasl[2,3], BBSc, MSc; Florian v Wangenheim[1], Dr; Tobias Kowatsch[1,4], PhD

[1]Department of Management, Technology, and Economics, Centre for Digital Health Interventions, ETH Zurich, Zurich, Switzerland

[2]Department of Educational Sciences, University of Potsdam, Potsdam, Germany

[3]International Max Planck Research School on the Life Course (LIFE), Berlin, Germany

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

[*]these authors contributed equally

Corresponding Author:
Christoph Gross, BSc, MA
Department of Management, Technology, and Economics
Centre for Digital Health Interventions
ETH Zurich
WEV J 409, Weinbergstr 56/58
Zurich, 8092
Switzerland
Phone: 41 78 779 70 9
Email: christophgross@ethz.ch

Abstract

Background: Recent years have witnessed a constant increase in the number of people with chronic conditions requiring ongoing medical support in their everyday lives. However, global health systems are not adequately equipped for this extraordinarily time-consuming and cost-intensive development. Here, conversational agents (CAs) can offer easily scalable and ubiquitous support. Moreover, different aspects of CAs have not yet been sufficiently investigated to fully exploit their potential. One such trait is the interaction style between patients and CAs. In human-to-human settings, the interaction style is an imperative part of the interaction between patients and physicians. Patient-physician interaction is recognized as a critical success factor for patient satisfaction, treatment adherence, and subsequent treatment outcomes. However, so far, it remains effectively unknown how different interaction styles can be implemented into CA interactions and whether these styles are recognizable by users.

Objective: The objective of this study was to develop an approach to reproducibly induce 2 specific interaction styles into CA-patient dialogs and subsequently test and validate them in a chronic health care context.

Methods: On the basis of the Roter Interaction Analysis System and iterative evaluations by scientific experts and medical health care professionals, we identified 10 communication components that characterize the 2 developed interaction styles: deliberative and paternalistic interaction styles. These communication components were used to develop 2 CA variations, each representing one of the 2 interaction styles. We assessed them in a web-based between-subject experiment. The participants were asked to put themselves in the position of a patient with chronic obstructive pulmonary disease. These participants were randomly assigned to interact with one of the 2 CAs and subsequently asked to identify the respective interaction style. Chi-square test was used to assess the correct identification of the CA-patient interaction style.

Results: A total of 88 individuals (42/88, 48% female; mean age 31.5 years, SD 10.1 years) fulfilled the inclusion criteria and participated in the web-based experiment. The participants in both the paternalistic and deliberative conditions correctly identified the underlying interaction styles of the CAs in more than 80% of the assessments ($X^2_{1,88}$=38.2; $P<.001$; phi coefficient r_ϕ=0.68). The validation of the procedure was hence successful.

Conclusions: We developed an approach that is tailored for a medical context to induce a paternalistic and deliberative interaction style into a written interaction between a patient and a CA. We successfully tested and validated the procedure in a web-based experiment involving 88 participants. Future research should implement and test this approach among actual patients with chronic diseases and compare the results in different medical conditions. This approach can further be used as a starting point to develop dynamic CAs that adapt their interaction styles to their users.

(J Med Internet Res 2021;23(1):e22919) doi:10.2196/22919

KEYWORDS

conversational agents; chatbots; human-computer interaction; physician-patient relationship; interaction styles, deliberative interaction; paternalistic interaction; digital health; chronic conditions; COPD

Introduction

Background

The interaction between patients and physicians is recognized as a critical success factor for treatment satisfaction, adherence, and subsequent treatment outcomes [1,2]. Its importance has been shown in face-to-face encounters between patients and physicians and in distance therapy via, for example, phone or internet [3,4]. A previous study [5] has differentiated between 4 distinct interaction styles between patients and physicians: paternalistic, informative, interpretative, and deliberative interaction styles. The paternalistic interaction style is characterized by physicians acting as “guardian” [5] of patients and paternally making decisions grounded in the assumption of objective and shared values between them and their patients. Although applying the informative interaction style, physicians act as “competent technical expert” [5], passing on information to their patients who ultimately have decision control based on their personal values. Within the interpretative interaction style, patients are confronted with medical information but are unsure about how to deal with them. Here, physicians act as a “counsellor or advisor” [5], helping patients to make their decision based on better self-understanding. Finally, physicians act as a “teacher or friend” [5] while engaging in a deliberative interaction style. They present medical information, promote particular health-related values, and conjointly discuss the best way forward together with their patients. This style can also be described as shared decision making and is advocated by contemporary medical research [6]. Especially while addressing patient autonomy in chronic care, a deliberative patient-physician interaction is thought to activate patients’ intrinsic values and goals better than any other interaction style and is, therefore, believed to be preferable [7].

However, such shared decision making is not always possible or even desired by patients. For example, patients seem to prefer paternalistic interactions in acute care conditions, especially when they have low health literacy or are emotionally overburdened by the situation [8]. Recent works have also shown that older people [9,10], men [9,10], less educated patients [10], patients with physical problems, and patients with severe exacerbation of their condition [11,12] prefer paternalistic interactions with their physicians. In addition, personal preferences for a preferred interaction style can change over time [13].

The personalization of physician-patient interactions thus seems to be appropriate while aiming at optimally adapting to the needs of the patients. This holds special importance in the context of chronic diseases, where patients have to deal with their condition for a prolonged period, experience functional limitations, require ongoing medical support, and often undergo several exacerbations of their condition [14-16]. In addition to continuing treatment and medical supervision of such chronic conditions, which pose an increasingly higher risk to the world population and are an enormous financial burden for global health care systems [17], it is important to engage in active disease management. Effective disease management includes educational measures, behavior modification, and psychological support [18] and can minimize overall harm and long-term effects of chronic conditions as well as confine exacerbation risks [19]. However, disease management is labor intensive, time-consuming, and costly for trained medical staff [20-23].

Against this background, emerging digital health tools such as conversational agents (CAs) offer hope in supporting patients’ self-management of their conditions. CAs are software programs that imitate natural interactions with human users by engaging in a human-like text-based and/or voice-enabled dialog [24,25]. Recent research has shown the ability of CAs to positively affect patient satisfaction [26], therapeutic alliance [27,28], and health-related outcomes [29,30]. Moreover, an overall acceptability has been established in various populations [31]. As scalable and ubiquitous digital tools, CAs facilitate personalized disease management outside the traditional health care system. Relevant aspects for developing personalized CAs, such as the required level of anthropomorphic appearance [32] or necessary design features [33], have been extensively investigated. However, to date, there has been no investigation of different interaction styles between patients and health care professionals where CAs play the role of medical experts. As mentioned above, these several forms of patient-physician interaction styles exert a significant effect on treatment success. We thus assume that appropriate interaction styles are also indispensable for patient-CA interactions. As a first step for developing personalized CAs that can adapt to the needs of the patient at hand and even adjust to changing individual preferences over time, it is essential to first develop and validate a systematic approach to develop and induce several interaction styles into patient-CA interactions.

Objectives

To this end, the objectives of this study were to (1) create and present a systematic approach to develop and induce specific interaction styles specifically for health care CAs and (2) validate whether individuals can correctly differentiate between the induced interaction styles. Our overarching research question for this paper is as follows: are humans capable of correctly identifying and labeling either an induced deliberative or paternalistic interaction style while interacting with a health care CA?

Methods

Development of CAs

In a 3-step process, we have developed a comprehensive mechanism to induce a deliberative and paternalistic interaction style into a CA interaction. These 3 steps are (1) development of the 2 interaction styles in the form of interaction items and a corresponding measurement scale, (2) scripting of CAs based on developed interaction styles, and (3) validation of the developed CAs in a web-based experiment.

We decided to induce deliberative and paternalistic interaction styles [5] as they are the 2 endpoints of a patient's autonomy spectrum [34]. Here, by means of education, patient autonomy and mutual trust increase from the paternalistic to the deliberative interaction style. Patient autonomy is understood as the ability to accept one's treatment preferences or to change to higher-order preferences through deliberation [34].

We first compiled an initial list of 28 communication components (Multimedia Appendix 1 [1,35-39]) that we adopted from the communication behaviors established by the Roter Interaction Analysis System (RIAS) [35]. The RIAS is a widely applied communication coding scheme for medical dialogs and assigns recorded verbal utterances to distinct categories [35]. Communication components combine communication clusters with communication categories to specify an explicit utterance in a dialog. To illustrate, consider the communication component "Therapeutic regimen_closed-ended question." It combines the communication cluster "Therapeutic regimen" with the communication category "closed-ended question" and denotes a remark by the physician about a patient's therapeutic regimen, posed as a closed-ended question. All communication components are described from the physician's point of view. On the basis of the related scientific literature [1,36,40-42], we assigned frequency levels (high or low) that are characteristic of a deliberative and paternalistic interaction style to each of the 28 communication components in the next step. For example, we allocated a high-frequency level of the communication component "Medical condition_Open-ended question" to the deliberative interaction style and a low frequency to the paternalistic one. Three health care practitioners of a European University Hospital have close experience with and exposure to both teaching and various medical communication techniques. They were thus qualified to review and endorse the authors' work.

To refine our proposed approach and improve its practicability, we aimed at reducing the total number of communication components. First, we excluded 6 communication components with identical frequency assignments for both the deliberative and paternalistic interaction styles, as they did not yield exclusive information on any of the 2 interaction strategies. An example would be "Biomedical information_About medical condition," where frequency levels were high for both the interaction styles. This initial triage resulted in a total of 22 remaining communication components. On the basis of our theoretical knowledge, we selected a binary choice between including and not including 15 of these components that seemed most relevant to distinguish between paternalistic and deliberative patient-physician-interactions. To triangulate our item selection and minimize the risk of bias, we asked 2 junior medical doctors with work experience less than 5 years since graduation and 2 experienced medical practitioners with more than 20 years of work experience to identify the 15 components they felt were the most relevant to differentiate between the 2 interaction styles. We deliberately chose to incorporate junior physicians because they are trained in the currently advocated deliberative interaction style and senior physicians because they are still familiar with the traditional paternalistic interaction style. The 4 physicians had different medical specializations. This approach allowed us to include both theoretical and practical perspectives while reducing the risk of bias in the item reduction process. To determine the intersecting set of communication components, we calculated the inter-rater reliability (IRR) across all the communication components [43]. Following this approach, we ensured the highest objectivity in the selection process. Consistent with a previous work [43], we chose 80% as the cut-off rate for the inclusion of an item. The evaluation ratings can be found in Multimedia Appendix 2. The final list resulted in 10 communication components, as listed in Table 1 (with example statements detailing the communication components) with their respective frequency measures for the 2 interaction styles. Among the final choices of the 10 communication components, 7 showed an IRR of 100%. Hence, all the 6 raters (ie, the main authors, TS and CG, and the 4 medical practitioners) selected these components to be the most relevant to differentiate between the 2 interaction styles.

Table 1. A final list of communication components.

ID	Communication component	Communication frequency		Example statement
		Paternalism	Deliberative	
CC1	Medical condition_open-ended question	Low	High	What can you tell me about the pain? [35]
CC2	Therapeutic regimen_open-ended question	Low	High	How are your symptoms developing since you take the new pills? (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])
CC3	Therapeutic regimen_closed-ended question	High	Low	Did you reduce your cigarette consumption to max. 10 cigarettes/day as discussed? (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])
CC4	Psychosocial Exchange about problems of daily living, issues about social relations, feelings, emotions	Low	High	It is important to talk about your worries regarding your condition (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])
CC5	Emotional Talk_Reassurance/Optimism	Low	High	Your arm will feel better soon, no worries! (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])
CC6	Emotional Talk_Empathy	Low	High	I can see how worried you are from hearing these results of your lung test. (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])
CC7	Emotional Talk_Partnership	Low	High	We'll get through this together [35]
CC8	Partnering and activation_Asking for patient opinion	Low	High	Do you want to bring your husband to the next session? (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])
CC9	Partnering and activation_Asking for understanding	Low	High	Can you follow all my instructions? (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])
CC10	Partnering and activation_Paraphrase and interpretation	Low	High	Ok, let me summarize what you told me about your symptoms; so you cough every night at least five times, you have constant pain in the left leg (developed by authors, adapted from the studies by Ong et al, 1995; Cavaco and Roter, 2010; Roter et al, 1997; and Cegala, 1997 [1,35-37])

The next step is dedicated to scripting the 2 CA interventions designed for patients with chronic obstructive pulmonary disease (COPD). We focused on COPD as a chronic condition, as the number of people affected by COPD continues to rise inexorably on a global scale, hence causing hardships among the affected and a tremendous financial burden on health care systems [44]. We based the content of the CA dialog on the teaching workbooks of "Living well with COPD," an evidence-based disease-specific self-management program originally developed at the Montreal Chest Institute in collaboration with the Respiratory Health Network of the Fonds de la Recherche en Santé du Quebec and Boehringer Ingelheim [45]. We decided to model the first day of this hypothetical patient-CA intervention as it incorporates relevant interaction categories (eg, introduction and patient education) of a comprehensive disease management intervention. In a two-step process, we first scripted a base-case CA interaction whose interaction style was as neutral as possible. In the second step, we followed our systematically developed specifications of communication components and corresponding frequency-level characteristics for the 2 interaction styles to develop 2 distinct CA scripts. This means that we reverse-engineered the labeling process of utterances depicted in the RIAS methodology to induce either a paternalistic or a deliberative interaction style into the base-case intervention, that is, editing, adding, or deleting separate sentences, parts of sentences, or terms to differentiate the frequency of communication items of the base-case CA interaction. We ensured that both the scripts were approximately of the same length, with 38 conversational turns (with 96 individual messages) in the deliberative and 32 (with 85 individual messages) in the paternalistic version. The 2 scripts differed in 40 instances, whereas this number included differences in the level of a single word as the smallest adjustable part of a sentence. These discrepancies are caused by the characteristics of the 2 interaction styles, and these discrepancies are to ensure the realism of the conversation flow. For instance, we defined that a deliberative interaction style has a high frequency of emotional talk around a partnership. In one of the first conversational turns, we thus scripted “Okay. Understood, then we are ready to start.” as a potential possibility of an answer in the deliberative script versus solely “Okay. Understood.” in the paternalistic script. The conversation tree had only one level of branch points and was then reverted to the main conversational flow. The average overall reading duration was 13.5 min for the deliberative and 12 min for the paternalistic version. The scripts were written in German. The 2 intervention variations were then presented to senior medical experts working in the pulmonary division of a European University Hospital, who assessed and confirmed realism. Excerpts of the scripts are shown and discussed in the Results section, and the complete scripts can be found in Multimedia Appendix 3. The intervention was purely text based without any visual or spoken cues to reduce any bias toward visual CA design features, such as gender, age, or visual appearance [33,46]. Hence, we named the CA Robo, a gender-neutral name.

Validation of CAs’ Interaction Styles

In the second phase, we validated the interaction styles and assessed whether participants engaging with 1 of the 2 CAs could identify the correct interaction style.

Experimental Design and Procedure

We conducted a closed, between-subjects web-based experiment, in which the participants were randomly prompted to engage with a CA that follows either a deliberative or paternalistic interaction style. Following the Checklist for Reporting of Results of Internet E-Surveys [47], we report on the design, procedure, and results of this experiment. Qualtrics software (Qualtrics XM), a software- and web-based survey and data collection platform, was used to design the experiment and to randomly assign participants to 1 of the 2 CAs. Collect.chat, a commercially available chatbot software, was used to develop chatbot dialogs. The CA was integrated into the Qualtrics HTML using an iframe. Before starting the web-based experiment, we tested its usability and technical functionality. The experiment was conducted between March 27 and April 11, 2020. The questionnaire comprised a total of 35 questions distributed over 8 pages (between 1 and 17 items per page). The respondents were not able to review and change their answers.

The experimental procedure was set up as follows: first, we informed the participants about the structure and length of the survey, its potential risks and confidentiality, data protection, and possible uses of the data. We also provided contact details of the investigators in the case of questions and comments. After receiving informed consent, we checked the participation conditions, such as being aged above 18 years and speakers of German language. We then queried a set of sociodemographic questions (age, gender, mother tongue, and education). Second, the participants were presented with a short and easy-to-comprehend scenario description that depicted the day of a patient with COPD who started using a CA (for details, refer to Multimedia Appendix 4), an established approach in health care for the investigation of specific aspects in a medical context [48,49]. The scenario prompted participants to put themselves in the position of a patient with COPD to be able to relate to the subsequent interaction with the CA [50,51]—a targeted health outcome was not included. We ensured that the necessary heterogeneity of participants was much better than that with a limited set of patients who are often homogeneous on the key characteristics such as age due to their shared medical condition (eg, COPD becomes clinically noticeable only from the age of 40-50 years [52]) by applying a scenario description with healthy participants instead of conducting the experiment with actual patients. Third, after presenting the scenario, the interaction with the CA started, which we embedded in a separate, dedicated page of the survey. The interaction was purely text based and comprised a prescripted dialog based on the 2 developed scripts, that is, one for the paternalistic and another for the deliberative interaction style. Participants chose between 1 to 3 predefined answer options. They interacted with the respective CA on their individual pace, with no supervision or guidance from the researchers. Fourth, after the interaction, the participants answered questions about the interaction style. Here, the participants were asked to choose which of the following 2 statements better described their perception of Robo’s interaction style with them: (1) Robo decides paternally, based on objective principles, or (2) Robo and I discuss and decide together. The former refers to a paternalistic, whereas the latter to a deliberative interaction style.

Participants

A priori power analysis was conducted using the R package (version 3.5.2) power analysis [53]. To identify a medium effect (r=0.30) in a chi-square test of independence at an α level of .05 and statistical power of 0.80, a total of 88 participants were required. Inclusion criteria of this study are as follows: participants had to be of aged above 18 years and German-speaking. The participants were recruited via email and social media through an anonymous link. The participants who were invited to participate in this study were from the authors’ academic institutions, networks, and cooperating partners, and participation was entirely voluntary, with no incentives offered. We intended to actively drive heterogeneity in the sample to ensure external validity. We recruited a total of 112 participants. Of these 112 participants, 24 did not complete the questionnaire and were thus excluded. Thus, the final sample comprised 88 persons (42/88, 48% female, mean age 31.5 years, SD 10.1 years), resulting in a completion rate of 79%. Table 2 shows an overview of the demographics of the participants. On average, the participants needed 29 min for the whole experiment, including the questionnaire and the CA interaction.

Table 2. Demographic characteristics of participants (n=88).

Characteristic	Value
Age in years, mean (SD)	31.5 (10.1)
Gender (female), n (%)	42 (48)
Education, n (%)	
<High school	15 (17)
High school	9 (10)
University degree	64 (73)

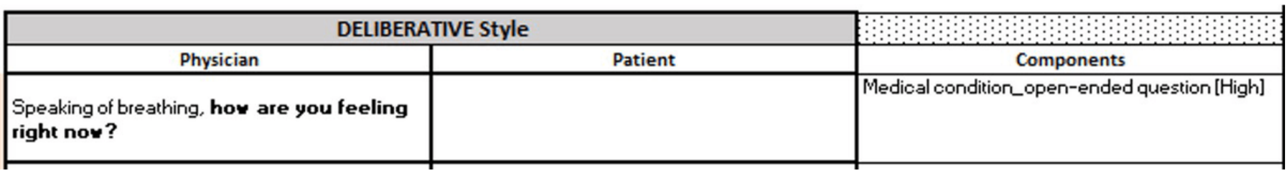
Data Analysis

To test whether the participants correctly identified the interaction type of the CAs, we applied a chi-square test of independence for a 2×2 contingency table and calculated the related phi coefficient using the statistical software R (version 3.5.2). The phi coefficient is a measure of the strength of the association between 2 binary variables. If it is positive, that is, the 2 variables are positively associated, then most of the data fall along the diagonal cells. To acknowledge the validation of the CAs as successful, we set 2 criteria based on the previous literature: (1) a significant chi-square test statistic with at least 80% of participants correctly identifying their assigned respective CA in each condition [43] and (2) a positive phi coefficient of at least a medium effect size ($r_\phi \geq 0.30$) [54].

Results

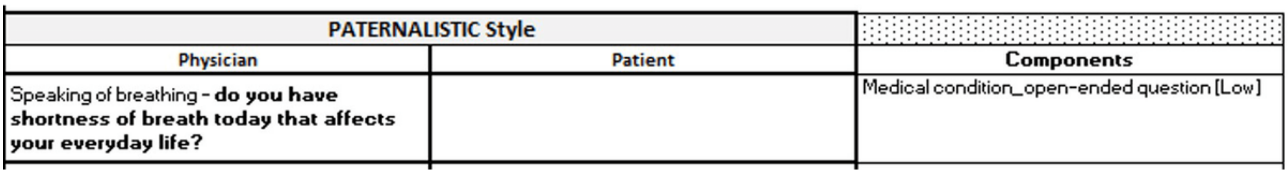
The results of the development of the deliberative and paternalistic interaction styles are presented in this section, whereas the development process of the interaction styles themselves is described in the Methods section. The described process resulted in 2 written scripts that were used for scripting the paternalistic and deliberative versions of the CA Robo. The scripts further included the answer options of the participants, the underlying communication items, and the applied communication frequency for each utterance. Figures 1 and 2 show an example of how the different interaction styles were induced into the CA script. The scripts were initially developed in German. This conversation has been translated into English by the authors for this study.

Figure 1. An example of the developed script for the deliberative version of the conversational agent (CA) Robo is shown here. The column labeled “Components” depicts the applied communication item and its respective communication frequency for the respective interaction style. The column labeled “Physician” describes what the CA Robo says, whereas the column “Patient” shows the reply options of the participant.



DELIBERATIVE Style		
Physician	Patient	Components
Speaking of breathing, **how are you feeling right now?**		Medical condition_open-ended question [High]

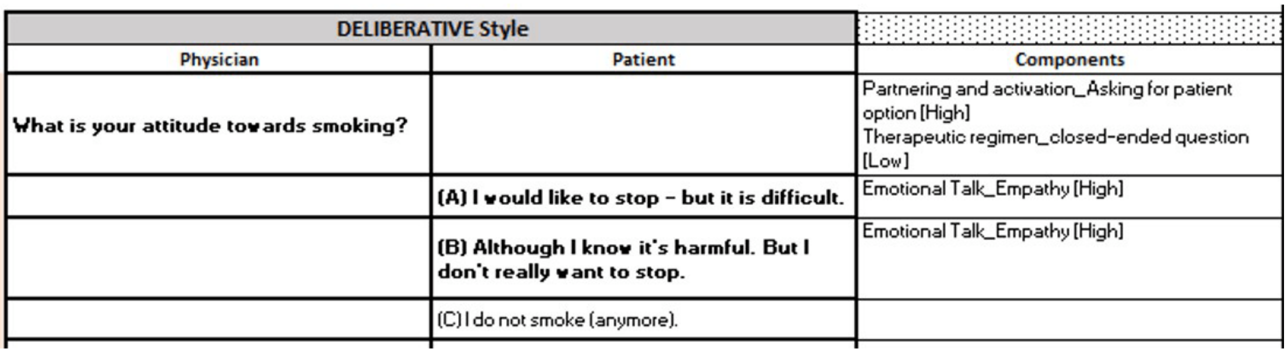
Figure 2. An example of the developed script for the paternalistic version of the conversational agent (CA) Robo is shown here. The column labeled “Components” depicts the applied communication item and its respective communication frequency for the respective interaction style. The column labeled “Physician” describes what the CA Robo says, whereas the column “Patient” shows the reply options of the participant.



PATERNALISTIC Style		
Physician	Patient	**Components**
Speaking of breathing - **do you have shortness of breath today that affects your everyday life?**		Medical condition_open-ended question [Low]

Figures 3 and 4 show an example of the interaction style scripts where the answers of the patients needed to deviate between the 2 interaction styles. This is because the interaction would have felt artificial when the same answer options would have been used for both the interaction styles.

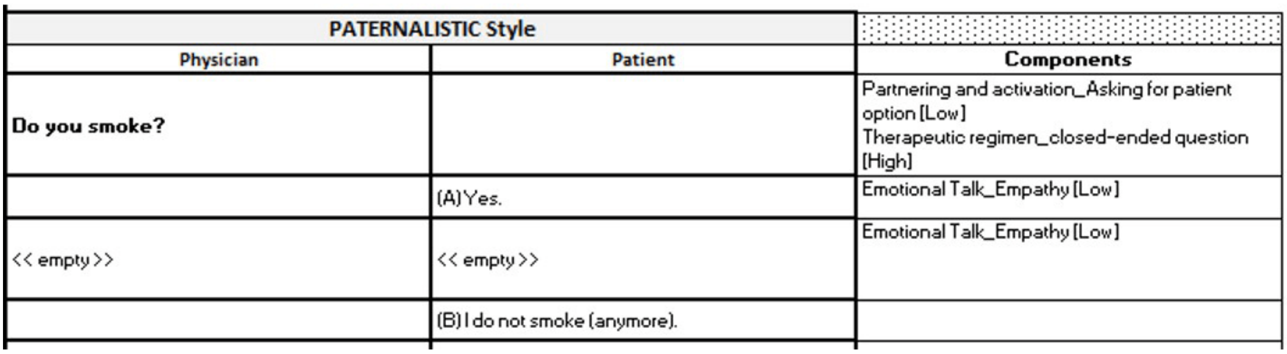
Figure 3. An example of the developed script for the deliberative version of the conversational agent Robo is shown here. Here, the answer options for patients, as depicted in the column "Patient," deviated between the 2 conditions.



DELIBERATIVE Style		
Physician	**Patient**	**Components**
What is your attitude towards smoking?		Partnering and activation_Asking for patient option [High] Therapeutic regimen_closed-ended question [Low]
	(A) I would like to stop – but it is difficult.	Emotional Talk_Empathy [High]
	(B) Although I know it's harmful. But I don't really want to stop.	Emotional Talk_Empathy [High]
	(C) I do not smoke (anymore).	

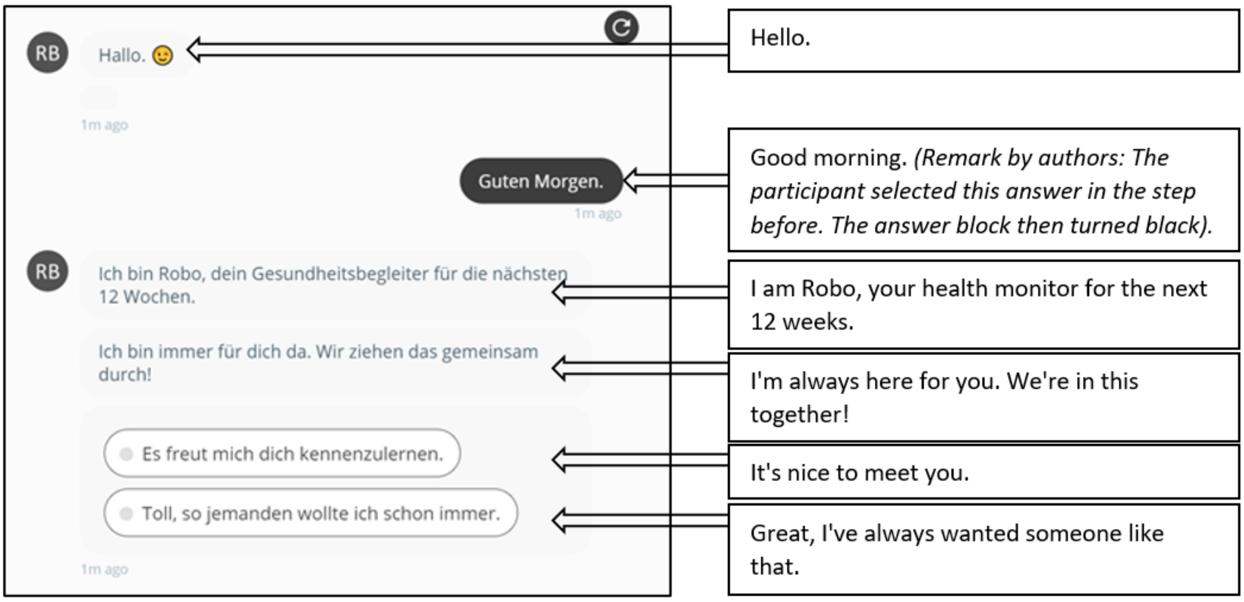
The developed scripts were implemented into the technical CA environment. In the web-based experiment, the participants could only see the utterances of the CA Robo (noted in the column "Physician" of Figures 1-4) and their possible answer options (noted in the column "Patient" of Figures 1-4). Figure 5 shows an exemplary conversational turn between a participant and the CA Robo as implemented in the web-based experiment.

Figure 4. An example of the developed script for the paternalistic version of the conversational agent Robo is shown here. Here, the answer options for the patients, as depicted in the column "Patient," deviated between the 2 conditions.



PATERNALISTIC Style		
Physician	**Patient**	**Components**
Do you smoke?		Partnering and activation_Asking for patient option [Low] Therapeutic regimen_closed-ended question [High]
	(A) Yes.	Emotional Talk_Empathy [Low]
<< empty >>	<< empty >>	Emotional Talk_Empathy [Low]
	(B) I do not smoke (anymore).	

Figure 5. An exemplary conversation snapshot between the conversational agent Robo and a participant. This figure depicts the start of the interaction, here applying the deliberative interaction style. The interaction was conducted in German; the authors added the English translations in the callouts for this study.



RB Hallo. 😊

1m ago

Guten Morgen.

1m ago

RB Ich bin Robo, dein Gesundheitsbegleiter für die nächsten 12 Wochen.

Ich bin immer für dich da. Wir ziehen das gemeinsam durch!

○ Es freut mich dich kennenzulernen.

○ Toll, so jemanden wollte ich schon immer.

1m ago

Hello.

Good morning. *(Remark by authors: The participant selected this answer in the step before. The answer block then turned black).*

I am Robo, your health monitor for the next 12 weeks.

I'm always here for you. We're in this together!

It's nice to meet you.

Great, I've always wanted someone like that.

We describe the results of the web-based experiment in the following section. On the basis of a 2×2 contingency table, a chi-square test of independence was performed to examine the relationship between the randomly assigned CA type and the correct identification of CA type by the participants (Table 3). The relationship between these variables was significant ($X^2_{1,88}$=38.2; P<.001), thereby indicating unequal frequency distributions between the cells. The associated phi coefficient of correlation was r_ϕ=0.68, corresponding to a large effect size [54]. Participants in both the paternalistic and deliberative conditions correctly identified their respective CA type more than 80% of the time: the probability of recognizing the paternalistic interaction style when it was, in fact, paternalistic was 37/(37+7)=84%. The same was true for the deliberative condition. Thus, based on the previously defined set of criteria (significant results of the chi-square test, 80% of participants correctly identified the respective CA in each condition, and phi coefficient $r_\phi \geq 0.30$), the validation of the interaction styles was successful.

Table 3. Contingency table of assigned and identified conversational agent interaction styles[a].

Identified by participants	Random assignment	
	P[b]	D[c]
P[b]	37	7
D[c]	7	37
	37/(37+7)=0.84	37/(37+7)=0.84

[a] 2x2 contingency table. The performance of 2 developed conversational agents was assessed by comparing the categories randomly assigned to participants by the experimenters with participants' own perception which category the assigned conversational agent belonged to. The type of conversational agent assigned by the experimenters is designated as paternalistic (P) or deliberative (D) and is listed above the 2x2 table.

[b] P: paternalistic.

[c] D: deliberative.

Discussion

Principal Findings

We developed an approach to induce deliberative and paternalistic interaction styles into a purely text-based patient-CA conversation [5]. It was developed for chronic health care applications against the background of the high relevance of chronic diseases and the patients' support potential of CAs. To our knowledge, this is the first attempt to develop and evaluate 2 different interaction styles for the interaction between a CA and a human user in a chronic health care context. This study successfully proves that humans can correctly identify and label an induced interaction style under experimental conditions. When randomly assigned, participants in both the paternalistic and deliberative experimental conditions correctly identified the interaction style in more than 80% of the cases. The procedure is based on modifying the frequency of communication items that are adopted from human-human medical interactions. Although we focused on the deliberative and paternalistic interaction styles, we expect the underlying methodology to be also applicable for inducing other interaction styles.

In the context of chronic diseases, a rising number of medical interventions is already based on the application of CAs [24,28,30,55-57]. However, studies investigating the interaction styles between humans and CAs are still scarce. Examples include a range of applications from health care to real estate and implicitly assume a default interaction between the agent and user [58-60]. In our opinion, this assumption of a standardized human-agent interaction disregards patients' preferences of different interaction styles as derived from human-human interactions. We addressed this shortcoming with our study; our approach allows the repeated and independent development of any patient-agent medical conversation with a deliberative and paternalistic interaction style. We therefore enable other researchers to develop and test these interaction styles in different health care contexts. This is the first but important building block to develop personalized CAs by better understanding and investigating which interaction styles offered by CAs are relevant in which health care situations (eg, acute vs chronic care situations or disease-specific situations), for which patient group (eg, for patient segmentation), and at which stage in the course of the disease and diseases-related dynamics (high- vs low-pain situations). By considering these additional factors in future work, we assume that the effectiveness of CAs can be further increased by deploying more personalized CA-based conversations that fit the specific medical context and the patient situation at hand.

As indicated above, the quality and effectiveness of the interaction between patients and physicians has already been established as decisive for treatment outcomes in human-human medical encounters [61,62]. This patient-physician interaction is acknowledged as a dynamic process in which personal preferences for the most effective communication can change over time [61,63]. Skilled physicians can adapt their interaction style to each patient and his or her situation, thereby improving the treatment outcome [63]. Until now, such dynamic and individual adaption of patient-physician interactions can only be realized within a human-human context. However, it would be beneficial to have digitized solutions that are capable of the same dynamic alterations in the form of adaptive CAs for providing on-demand individualized medical support. Such solutions would yield hope for increased treatment adherence and subsequently improved medical outcomes, especially within the context of chronic diseases that often require prolonged medical oversight and support. The first part of developing such dynamic interactive CAs is to build their ability to provide more than one interaction style for communication with human users. In this paper, we provided this first stage by showing that 2 different interaction styles can be successfully developed, implemented into a CA-patient communication, and correctly identified and labeled by humans in the context of chronic disease health care.

Strengths and Limitations

This study has several strengths. By adapting the RIAS [35], a validated and widely used coding scheme for medical dialogs, we ensured an objective approach to identify the key communication items that induce the interaction styles. We detailed and refined the procedure by integrating the knowledge of novice and experienced medical experts, thereby ensuring a broad range of expertise in terms of patient-physician interaction behavior taught in medical school as well as practical experience with patients. We further measured the IRR to objectively calculate the agreement between the experts and our assessment and to select the most important items. Moreover, we tested the developed scripts with additional medical experts before scripting the CAs used in the web-based experiment.

A limitation of this study is that participants were all German-speaking and based in Switzerland, Germany, and Austria. Other languages might have different requirements regarding essential communication items for representing a deliberative and paternalistic interaction style. Cultural differences with respect to power distance, individualism, and uncertainty avoidance, as analyzed by Hofstede [64], might also play a significant role. Another limitation is the focus on one specific chronic condition. Although we consider our methodology sufficient to model medical conversations for other disease conditions, we only tested it in the context of COPD. Patients with other conditions—acute or chronic—might have other demands for their interaction with a digital CA. In addition, the CA in our experiment was rule based with prescripted answer possibilities. Although this was necessary to control the experimental conditions between the 2 interaction conditions, there is a rising number of artificial intelligence (AI)–based CAs for health care apps. These agents do not rely on prescripted conversations but often allow for natural interaction using unconstrained written, spoken, or visual input [24,25,56]. The applicability of our procedure for AI-based agents should be evaluated in future research.

Suggestions for Future Research

In general, we advise future research in the field of digital health care to put a stronger focus on the consideration of different interaction styles between human users and CAs. We propose that such CAs are based on our approach for inducing either a paternalistic or deliberative interaction style. It still must be determined which of the 2 interaction styles is the best situational fit for an individual engaging with a CA. In the following step, dynamic CAs can adapt their interaction style to both the personal and situational circumstances of individual users, much like human physicians are already able to do [61,63]. We suggest evaluating the implementation of human-agent interaction styles in different medical contexts, such as various acute and chronic conditions, as well as with different health care goals such as diagnosis, treatment, or patient education.

Furthermore, CAs with varying interaction styles should be developed in various languages and in different cultures. This would allow the investigation of the role of language and culture in interaction preferences. In addition, there might be more than the described 2 interaction styles that are of interest in the medical context. For example, Emanuel and Emanuel [5] also described an informative and interpretative interaction style. These could also be adapted for and integrated into CAs to develop a broader range of options for patient-CA interactions, for example, in the context of the information about invasive examination or interventions.

Moreover, we only modeled and tested the first day of an interaction between a patient and a CA. It would be interesting to see the effect of longer interventions. This is especially relevant for chronic conditions, as they require ongoing treatment and accompanying digital interventions for potentially the entire treatment period.

Furthermore, we suggest the implementation of our approach to the architecture of an AI-based CA to evaluate its applicability for this type of technology. Finally, because our CA was purely text based, it would be interesting to see what role other communication forms such as visualizations, embodiment, or spoken interaction play for the induction of particular interaction styles.

Conflicts of Interest

FvW co-chairs the Center for Digital Health Interventions (CDHI), a joint initiative between the Department of Management, Technology and Economics at Eidgenössische Technische Hochschule Zürich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. TK is the Scientific Director of the CDHI. He is also a co-founder of Pathmate Technologies, a university spin-off company that delivers digital clinical pathways with the help of CAs. However, Pathmate Technologies is not involved in the study described in this paper. CG, TS, and AH report no conflicts of interest.

Multimedia Appendix 1
Overview items.
[PDF File (Adobe PDF File), 305 KB - jmir_v23i1e22919_app1.pdf]

Multimedia Appendix 2
Inter-rater reliability item selection.
[PDF File (Adobe PDF File), 131 KB - jmir_v23i1e22919_app2.pdf]

Multimedia Appendix 3
Scripts of conversational agents in German and English.
[PDF File (Adobe PDF File), 229 KB - jmir_v23i1e22919_app3.pdf]

Multimedia Appendix 4
Scenario description in German and English.
[PDF File (Adobe PDF File), 78 KB - jmir_v23i1e22919_app4.pdf]

References

1. Ong L, de Haes J, Hoos A, Lammes F. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [doi: 10.1016/0277-9536(94)00155-m] [Medline: 7792630]
2. Kaplan SH, Greenfield S, Ware JE. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. Med Care 1989 Mar;27(3 Suppl):S110-S127. [doi: 10.1097/00005650-198903001-00010] [Medline: 2646486]
3. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
4. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet 2001 Mar 10;357(9258):757-762. [doi: 10.1016/s0140-6736(00)04169-6] [Medline: 11253970]
5. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. J Am Med Assoc 1992 Apr;267(16):2221-2226. [doi: 10.1001/jama.1992.03480160079038] [Medline: 1556799]
6. Heritage J, Maynard DW. Problems and prospects in the study of physician-patient interaction: 30 years of research. Annu Rev Sociol 2006 Aug;32(1):351-374. [doi: 10.1146/annurev.soc.32.082905.093959]
7. Borza LR, Gavrilovici C, Stockman R. Ethical models of physician--patient relationship revisited with regard to patient autonomy, values and patient education. Rev Med Chir Soc Med Nat Iasi 2015;119(2):496-501. [Medline: 26204658]
8. Montori VM, Gafni A, Charles C. A shared treatment decision-making approach between patients with chronic conditions and their clinicians: the case of diabetes. Health Expect 2006 Mar;9(1):25-36 [FREE Full text] [doi: 10.1111/j.1369-7625.2006.00359.x] [Medline: 16436159]
9. Blanchard CG, Labrecque MS, Ruckdeschel JC, Blanchard EB. Information and decision-making preferences of hospitalized adult cancer patients. Soc Sci Med 1988 Jan;27(11):1139-1145. [doi: 10.1016/0277-9536(88)90343-7] [Medline: 3206248]
10. Benbassat J, Pilpel D, Tidhar M. Patients' preferences for participation in clinical decision making: a review of published surveys. Behav Med 1998 Jan;24(2):81-88. [doi: 10.1080/08964289809596384] [Medline: 9695899]
11. Say R, Murtagh M, Thomson R. Patients' preference for involvement in medical decision making: a narrative review. Patient Educ Couns 2006 Mar;60(2):102-114. [doi: 10.1016/j.pec.2005.02.003] [Medline: 16442453]
12. Gibson PG, Talbot PI, Toneguzzi RC. Self-management, autonomy, and quality of life in asthma. Population Medicine Group 91C. Chest 1995 Apr;107(4):1003-1008. [doi: 10.1378/chest.107.4.1003] [Medline: 7705105]
13. Strull WM. Do patients want to participate in medical decision making? J Am Med Assoc 1984 Dec 07;252(21):2990. [doi: 10.1001/jama.1984.03350210038026] [Medline: 6502860]
14. Hvidberg MF, Johnsen SP, Glümer C, Petersen KD, Olesen AV, Ehlers L. Catalog of 199 register-based definitions of chronic conditions. Scand J Public Health 2016 Jul 20;44(5):462-479 [FREE Full text] [doi: 10.1177/1403494816641553] [Medline: 27098981]
15. VAN DEN BOS GA. The burden of chronic diseases in terms of disability, use of health care and healthy life expectancies. Eur J Public Health 1995;5(1):29-34. [doi: 10.1093/eurpub/5.1.29]
16. Han MK, Quibrera PM, Carretta EE, Barr RG, Bleecker ER, Bowler RP, et al. Frequency of exacerbations in patients with chronic obstructive pulmonary disease: an analysis of the SPIROMICS cohort. Lancet Respir Med 2017 Aug;5(8):619-626. [doi: 10.1183/13993003.congress-2016.pa676] [Medline: 28668356]
17. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar 10;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
18. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4. [Medline: 10345255]
19. Frey U, Suki B. Complexity of chronic asthma and chronic obstructive pulmonary disease: implications for risk assessment, and disease progression and control. Lancet 2008 Sep 20;372(9643):1088-1099 [FREE Full text] [doi: 10.1016/S0140-6736(08)61450-6] [Medline: 18805337]
20. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Joint National Committee on Prevention, Detection, Evaluation,Treatment of High Blood Pressure. National Heart, Lung,Blood Institute, National High Blood Pressure Education Program Coordinating Committee. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension 2003 Dec;42(6):1206-1252. [doi: 10.1161/01.HYP.0000107251.49515.c2] [Medline: 14656957]

21. Stone NJ. Nonpharmacologic management of mixed dyslipidemia associated with diabetes mellitus and the metabolic syndrome: a review of the evidence. Am J Cardiol 2008 Dec 22;102(12A):14L-18L. [doi: 10.1016/j.amjcard.2008.09.069] [Medline: 19084085]
22. van der Wal MH, van Veldhuisen DJ, Veeger NJ, Rutten FH, Jaarsma T. Compliance with non-pharmacological recommendations and outcome in heart failure patients. Eur Heart J 2010 Jun 30;31(12):1486-1493. [doi: 10.1093/eurheartj/ehq091] [Medline: 20436049]
23. Wise RA, Tashkin DP. Optimizing treatment of chronic obstructive pulmonary disease: an assessment of current therapies. Am J Med 2007 Aug;120(8 Suppl 1):S4-13. [doi: 10.1016/j.amjmed.2007.04.007] [Medline: 17678942]
24. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
25. Schachner T, Keller R, Wangenheim FV. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
26. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
27. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 2005 Jun;12(2):293-327. [doi: 10.1145/1067860.1067867]
28. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
29. Fadhil A, Gabrielli S, Kessler F. Addressing challenges in promoting healthy lifestyles: the Al-chatbot approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May, 2017; Barcelona Spain p. 261-265. [doi: 10.1145/3154862.3154914]
30. Ma T, Sharifi H, Chattopadhyay D. Virtual Humans in Health-Related Interventions: A Meta-Analysis. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May, 2019; Glasgow Scotland Uk p. 1-6. [doi: 10.1145/3290607.3312853]
31. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
32. Kang SH, Feng AW, Leuski A, Casas D, Shapiro A. The effect of an animated virtual character on mobile chat interactions. In: Proceedings of the 3rd International Conference on Human-Agent Interaction. 2015 Presented at: HAI 2015: The Third International Conference on Human-Agent Interaction; October, 2015; Daegu Kyungpook Republic of Korea p. 105-112.
33. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: a literature review. Int J Hum Comput Stud 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
34. Reach G. Patient autonomy in chronic care: solving a paradox. Patient Prefer Adherence 2013 Dec 12;8:15-24 [FREE Full text] [doi: 10.2147/PPA.S55022] [Medline: 24376345]
35. Cavaco A, Roter D. Pharmaceutical consultations in community pharmacies: utility of the Roter Interaction Analysis System to study pharmacist-patient communication. Int J Pharm Pract 2010 Jun;18(3):141-148. [Medline: 20509347]
36. Roter DL, Stewart M, Putnam SM, Lipkin M, Stiles W, Inui TS. Communication patterns of primary care physicians. J Am Med Assoc 1997;277(4):350-356. [Medline: 9002500]
37. Cegala DJ. A study of doctors' and patients' communication during a primary care consultation: implications for communication training. J Health Commun 1997;2(3):169-194. [doi: 10.1080/108107397127743] [Medline: 10977246]
38. Ong LM, Visser MR, Kruyver IP, Bensing JM, van den Brink-Muinen A, Stouthard JM, et al. The Roter Interaction Analysis System (RIAS) in oncological consultations: psychometric properties. Psychooncology 1998;7(5):387-401. [doi: 10.1002/(SICI)1099-1611(1998090)7:5<387::AID-PON316>3.0.CO;2-G] [Medline: 9809330]
39. Detmar SB, Muller MJ, Wever LD, Schornagel JH, Aaronson NK. The patient-physician relationship. Patient-physician communication during outpatient palliative treatment visits: an observational study. J Am Med Assoc 2001 Mar 14;285(10):1351-1357. [doi: 10.1001/jama.285.10.1351] [Medline: 11255393]
40. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3] [Medline: 9032835]
41. Street RL. Communicative styles and adaptations in physician-parent consultations. Soc Sci Med 1992 May;34(10):1155-1163. [doi: 10.1016/0277-9536(92)90289-3] [Medline: 1641677]
42. Williams S, Weinman J, Dale J. Doctor-patient communication and patient satisfaction: a review. Fam Pract 1998 Oct;15(5):480-492. [doi: 10.1093/fampra/15.5.480] [Medline: 9848436]
43. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]

44. Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. Lancet 2007 Sep 1;370(9589):765-773. [doi: 10.1016/S0140-6736(07)61380-4] [Medline: 17765526]
45. Bourbeau J, Julien M, Maltais F, Rouleau M, Beaupré A, Bégin R, Chronic Obstructive Pulmonary Disease axis of the Respiratory Network Fonds de la Recherche en Santé du Québec. Reduction of hospital utilization in patients with chronic obstructive pulmonary disease: a disease-specific self-management intervention. Arch Intern Med 2003 Mar 10;163(5):585-591. [doi: 10.1001/archinte.163.5.585] [Medline: 12622605]
46. McDonnell M, Baxter D. Chatbots and gender stereotyping. Interact Comput 2019 Mar;31(2):116-121. [doi: 10.1093/iwc/iwz007]
47. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
48. Offredy M, Kendall S, Goodman C. The use of cognitive continuum theory and patient scenarios to explore nurse prescribers' pharmacological knowledge and decision-making. Int J Nurs Stud 2008 Jun;45(6):855-868. [doi: 10.1016/j.ijnurstu.2007.01.014] [Medline: 17362959]
49. Offredy M. Decision-making in primary care: outcomes from a study using patient scenarios. J Adv Nurs 2002 Dec;40(5):532-541. [doi: 10.1046/j.1365-2648.2002.02410.x] [Medline: 12437602]
50. Bahari SF. Qualitative versus quantitative research strategies: contrasting epistemological and ontological assumptions. Jurnal Teknologi 2012 Jan 20:17-28. [doi: 10.11113/jt.v52.134]
51. Bell E, Bryman A, Harley B. Business research methods. In: Research Methods. New York: Oxford University press; 2018.
52. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. Lancet 2015 Mar 07;385(9971):899-909. [doi: 10.1016/S0140-6736(14)60446-3] [Medline: 25123778]
53. Fan FY. Power Analysis in Experimental Design. cran.R-Project. 2017. URL: https://cran.r-project.org/web/packages/powerAnalysis/powerAnalysis.pdf [accessed 2020-12-24]
54. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: L Erlbaum Associates; 1988.
55. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
56. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
57. Kramer J, Künzler F, Mishra V, Smith S, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
58. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
59. Cassell J, Bickmore T. Negotiated collusion: modeling social language and its relationship effects in intelligent agents. User Model User-adapt Interact 2003 Feb;13(1-2):89-132. [doi: 10.1023/A:1024026532471]
60. Kidd C. Designing for Long-Term Human-Robot Interaction and Application to Weight Loss. Program in Media Arts and Sciences, School of Architecture and Planning at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY. 2008. URL: https://dspace.mit.edu/handle/1721.1/42407 [accessed 2021-01-12]
61. Angus D, Watson B, Smith A, Gallois C, Wiles J. Visualising conversation structure across time: insights into effective doctor-patient consultations. PLoS One 2012;7(6):e38014 [FREE Full text] [doi: 10.1371/journal.pone.0038014] [Medline: 22693629]
62. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
63. Loughlan C, Mutrie N. Conducting an exercise consultation: guidelines for health professionals. J Inst Heal Educ 2014 Mar 10;33(3):78-82. [doi: 10.1080/03073289.1996.10805880]
64. Hofstede G. National cultures in four dimensions: a research-based theory of cultural differences among nations. Int Stud Manag Organ 2016 Feb 02;13(1-2):46-74. [doi: 10.1080/00208825.1983.11656358]

Abbreviations

AI: artificial intelligence
CA: conversational agent
COPD: chronic obstructive pulmonary disease
IRR: inter-rater reliability
RIAS: Roter Interaction Analysis System

Edited by G Eysenbach; submitted 27.07.20; peer-reviewed by E Bellei, A Kocaballi; comments to author 24.09.20; revised version received 11.11.20; accepted 07.12.20; published 29.01.21.

Please cite as:
Schachner T, Gross C, Hasl A, v Wangenheim F, Kowatsch T
Deliberative and Paternalistic Interaction Styles for Conversational Agents in Digital Health: Procedure and Validation Through a Web-Based Experiment
J Med Internet Res 2021;23(1):e22919
URL: http://www.jmir.org/2021/1/e22919/
doi:10.2196/22919
PMID:33512328

Original Paper

Economic Burden and Health Care Access for Patients With Inflammatory Bowel Diseases in China: Web-Based Survey Study

Qiao Yu[1*], MD, PhD; Chunpeng Zhu[1*], MD, PhD; Shuyi Feng[1], MD; Liyi Xu[1], MD; Shurong Hu[1], MD, PhD; Hao Chen[1], MD; Hanwen Chen[1], MD; Sheng Yao[1], MD; Xiaoying Wang[1], MD; Yan Chen[1], MD, PhD

Department of Gastroenterology, the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Yan Chen, MD, PhD
Department of Gastroenterology, the Second Affiliated Hospital, Zhejiang University School of Medicine
No 88, Jiefang Road
Hangzhou, 310000
China
Phone: 86 13757118653
Email: chenyan72_72@zju.edu.cn

Abstract

Background: The increasing incidence of inflammatory bowel disease (IBD) has imposed heavy financial burdens for Chinese patients; however, data about their financial status and access to health care are still lacking. This information is important for informing patients with IBD about disease treatment budgets and health care strategies.

Objective: The aim of this study was to evaluate the economic status and medical care access of patients with IBD through the China Crohn's & Colitis Foundation web-based platform in China.

Methods: Our study was performed in 14 IBD centers in mainland China between 2018 and 2019 through WeChat. Participants were asked to complete a 64-item web-based questionnaire. Data were collected by the Wenjuanxing survey program. We mainly focused on income and insurance status, medical costs, and access to health care providers. Respondents were stratified by income and the associations of income with medical costs and emergency visit times were analyzed.

Results: In this study, 3000 patients with IBD, that is, 1922 patients with Crohn disease, 973 patients with ulcerative colitis, and 105 patients with undetermined colitis were included. During the last 12 months, the mean (SD) direct and indirect costs for per patient with IBD were approximately US $11,668.68 ($7944.44) and US $74.90 ($253.60) in China. The average reimbursement ratios for most outpatient and inpatient costs were less than 50%. However, the income of 85.5% (2565/3000) of the patients was less than ¥10,000 (US $1445) per month. Approximately 96.5% (2894/3000) of the patients were covered by health insurance, but only 24.7% (741/3000) of the patients had private commercial insurance, which has higher imbursement ratios. Nearly 98.0% (2954/3000) of the patients worried about their financial situation. Thus, 79.7% (2392/3000) of the patients with IBD tried to save money for health care and even delayed their medical treatments. About half of the respondents (1282/3000, 42.7%) had no primary care provider, and 52.2% (1567/3000) of the patients had to visit the emergency room 1-4 times per year for the treatment of their IBD. Multivariate analysis revealed that lower income (P=.001) and higher transportation (P=.004) and accommodation costs (P=.001) were significantly associated with the increased number of emergency visits of the patients.

Conclusions: Chinese patients with IBD have enormous financial burdens and difficulties in accessing health care, which have increased their financial anxiety and inevitably influenced their disease outcomes. Early purchase of private insurance, thereby increasing the reimbursement ratio for medical expenses, and developing the use of telemedicine would be effective strategies for saving on health care costs.

(J Med Internet Res 2021;23(1):e20629) doi:10.2196/20629

KEYWORDS

inflammatory bowel disease; Crohn disease; ulcerative colitis; primary care provider; emergency room; eHealth; gastroenterology; proctology

Introduction

Inflammatory bowel disease (IBD) is a group of disorders that cause sections of the gastrointestinal tract to become inflamed and ulcerated. IBD imposes a significant impact on the quality of life through ongoing symptoms, including reduced ability to work, social stigma, and restriction in career choices. IBD causes a great burden globally because of the direct costs of care and the indirect costs associated with disability and missed work [1-4]. In the United States, IBD ranks as 1 of the 5 most expensive gastrointestinal disorders despite it being the lowest in prevalence in the list of gastrointestinal disorders [5]. The total financial burden of IBD in the United States was estimated to be US $14.6 billion to US $31.6 billion in 2014 [6]; however, recent data have indicated that the total costs may far exceed these earlier estimates [7]. In Canada, wherein the prevalence of IBD is one of the highest in the world, the economic cost for IBD was conservatively estimated to be over US $0.9 billion in 2018 [8]. In Europe, 2.5-3 million people have been estimated to be affected with IBD with a direct health care cost of 4.6-5.6 billion Euros/year [9]. However, the treatment for IBD is not curative. Clinical management aims at inducing and maintaining remission by using pharmaceutical agents and surgery, and evolving treatment guidelines advocate rapid scale-up to biological agents for improving health outcomes and quality of life. Consequently, the health care costs are driven by the increasing use of biological agents—most importantly by anti-tumor necrosis factor-α therapy [10]. In China, retrospective analysis has shown an increasing incidence of IBD; it is estimated that the IBD prevalence in China, which includes up to 11.6 ulcerative colitis cases per 100,000 person-years and 1.4 Crohn disease cases per 100,000 person-years [11,12], is the highest in Asia [13].

China is facing a growing burden with the increased use of health care resources, including outpatient visits, emergency room visits, hospital admissions, and surgeries [14]. Medical insurance policies vary widely among provinces in China. It is important to understand the financial burden and health care access of Chinese patients with IBD to measure treatment values and optimize health care policies. However, the collection of accurate epidemiologic data in China has been hampered by the lack of a nationwide IBD registry. One study from China in 2017 reported that 30.6% of the patients with IBD spent over half of their income to cover medical costs [15]. A retrospective cohort study in Hongkong reported that the total direct medical expenditure was US $7,072,710 for 435 patients with IBD, wherein hospitalizations (33%) and 5-aminosalicylic acid (23%) therapy accounted for the most part [16]. However, economic data and association between financial burden and health care access in mainland China are still lacking.

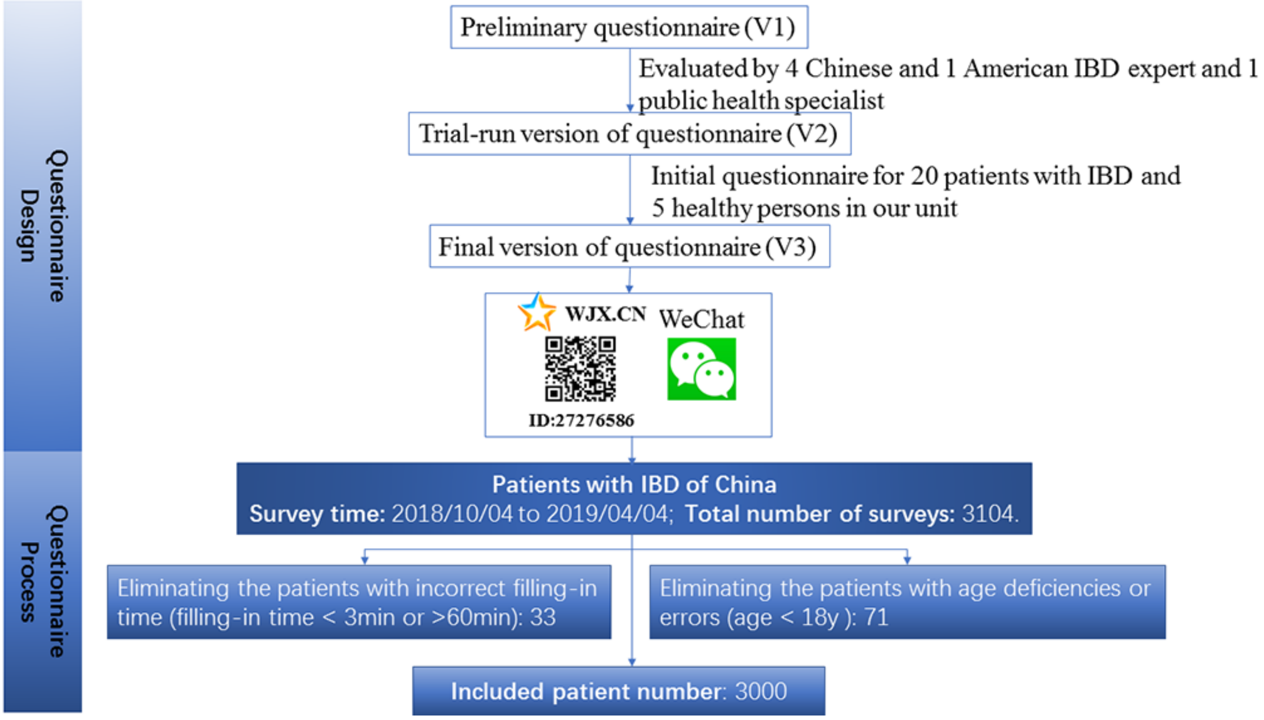
The China Crohn’s & Colitis Foundation (CCCF) serves as a nonprofit, volunteer-driven organization dedicated to improving the quality of life of patients with IBD in China [14]. With the support of the CCCF platform, we sought to provide a national snapshot of the current status of care for patients with IBD through electronic questionnaires in 14 IBD centers in mainland China. The contents of our survey focused on health care expenditures and insurance, access to care and therapies, affordability, and financial stress related to IBD.

Methods

Survey Development and Delivery

We used a 64-item questionnaire developed by Rubin et al [1], which was modified from the IBD questionnaire and guidelines for the comprehensive intervention of chronic diseases in China to assess the financial burdens and health care access of patients with IBD in China. Our questionnaire focused on topics, including respondent disease type, disease severity, disease duration, IBD-related treatment, access to provider care, employment and insurance status, income, and strategies for affording care. As shown in Figure 1, the questionnaire was evaluated and modified by CCCF doctors from 14 IBD centers in mainland China and was produced by Wenjuanxing [17], which is a free and open platform for survey design. We delivered the web-based survey from October 4, 2018 to April 4, 2019 through the CCCF’s WeChat public platform, which is a popular source for patient education about IBD in China [18]. The inclusion criteria were informed consent, older than 18 years, and diagnosed with Crohn disease, ulcerative colitis, or indeterminate colitis. The questionnaires without age information were excluded. This study was approved by the medical ethics committee of the Second Affiliated Hospital, School of Medicine of Zhejiang University (No. 314).

Figure 1. Flow chart depicting the selection of the survey sample for analysis.



Clinical Variables

Demographic variables included IBD type (Crohn disease, ulcerative colitis, or indeterminate colitis), marital and work status, education level, income, and insurance status. Direct costs in this study mainly included outpatient costs, hospitalization costs, biological agent costs, and surgery costs. As biological agent costs were partly repeated with hospitalization costs, it was not calculated into the direct costs here. Indirect costs were defined as nonprescription medications, transportation costs, accommodation costs, etc; however, we only calculated the transportation and accommodation costs in our study. The clinical variables of indirect costs included transportation and accommodation costs. Variables of health care access included emergency room visit frequency, hospitalization duration, and waiting time for hospitalization.

Statistical Analyses

Statistical analyses were conducted using SPSS software (version 25.0, IBM Corp). To estimate the average expenses, the mean and variance were calculated from the median value of the finite interval and the minimum value of the infinite interval. A binary logistic regression model was used to analyze the association between income or costs and variables of health care access.

Results

Population Characteristics

As shown in Figure 1, 3104 individuals from 14 IBD centers completed the questionnaire. Of those, 3000 respondents with a diagnosis of IBD were included for the analysis. Participants were eliminated from the data if they were younger than 18 years or did not provide age information (n=71) or if they completed the survey in less than 3 minutes or more than 60 minutes (n=33). The patients with IBD who completed the survey were distributed throughout all regions of mainland China. Most respondents were patients with Crohn disease (1922/3000, 64.1%), followed by those with ulcerative colitis (973/3000, 32.4%) or unclassified IBD (105/3000, 3.5%). Most respondents reported a disease duration of IBD of less than 5 years (2486/3000, 89.3%). The median age of the respondents was 34 years (range 18-73 years). Approximately 68.1% (2042/3000) of the patients were married (Table 1). The employment status of the patients with IBD was also assessed in our survey; 36.4% (1789/3000) of the patients had full-time work, 18.5% (555/3000) had part-time work, 6.9% (207/3000) were temporarily unemployed due to IBD, 20.7% (621/3000) were under the working age, 6.5% (196/3000) reported nonparticipation attributable to IBD, 6.7% (201/3000) were out of work, and 4.2% (127/3000) were retired.

Table 1. Background characteristics of the patients with inflammatory bowel disease (n=3000).

Characteristics	Value
Sex, n (%)	
Female	1211 (40.4)
Male	1789 (59.6)
Inflammatory bowel disease type, n (%)	
Crohn disease	1922 (64.1)
Ulcerative colitis	973 (32.4)
Inflammatory bowel disease (unclassified)	105 (3.5)
Inflammatory bowel disease duration (years), n (%)	
0-5	2486 (89.3)
>5	514 (10.7)
Median age (range)	34 (18-73)
Marital status, n (%)	
Married	2042 (68.1)
Unmarried	958 (31.9)
Work status, n (%)	
Working full-time	1093 (36.4)
Working part-time	555 (18.5)
Nonparticipation attributable to inflammatory bowel disease	207 (6.9)
Underage	621 (20.7)
Out of work	196 (6.5)
Currently unemployed	201 (6.7)
Retired	127 (4.2)
Highest level of education attained, n (%)	
Illiterate	17 (0.6)
Elementary/Junior/Senior school	1389 (46.3)
College degree	1473 (49.1)
Graduated or higher level education	121 (4.0)
Income[a] per month, n (%)	
<¥5000 (<US $723)	1490 (49.7)
¥5000-¥9999 (US $723-$1445)	1073 (35.8)
≥¥10,000 (≥US $1445)	437 (14.5)
Insurance status, n (%)	
Uninsured	106 (3.5)
Urban medical insurance	1568 (52.3)
New type of rural cooperative medical care	1011 (33.7)
Student health insurance	157 (5.2)
Other medical insurance	158 (5.3)
Commercial insurance, n (%)	
Yes	741 (24.7)
No	2259 (75.3)
Primary care provider, n (%)	
No primary care provider	1282 (42.7)

Characteristics	Value
Primary care provider	1718 (57.3)
Inflammatory bowel disease expert[b]	1182 (68.8)

[a] US $1=¥6.9197.

[b] As identified by the patients.

Financial Burden of IBD

The financial burden of IBD includes the direct and indirect costs incurred by individuals and society beyond the health care system. In estimation, the mean (SD) direct and indirect costs for per patient with IBD per year were conservatively ¥80,743.73 (¥54,973.12) (US $11,668.68 [$7944.44]) and ¥518.27 (¥1754.85) (US $74.90 [$253.60]), respectively. As shown in Table 2, for total outpatient costs, 41.7% (1252/3000) of the respondents with IBD reported spending more than ¥20,000 (US $2890) during the last 12 months. However, 46.5% of the patients (1394/3000) had no reimbursement for the outpatient costs. With respect to inpatient costs, 69.3% (2080/3000) of the patients with IBD reported paying less than ¥50,000 (US $7226) for hospitalizations during the last 12 months, while 30.7% (920/3000) of them paid more than ¥50,000 (US $7226) for their hospitalizations. The proportion of reimbursement for the hospitalization costs was more than 50% for 54.6% (1171/2143) of the patients with IBD. Among all the respondents, only 26.6% (799/3000) of the patients with IBD received more than 3 infusions of biological agents per year. However, 61.7% (626/1014) of them covered the costs of the biological agents at their own expense without any reimbursement. Moreover, 23.6% (707/3000) of the patients with IBD reported experiencing at least one surgery throughout their disease duration.

Table 2. Health care costs for inflammatory bowel disease treatment during the last 12 months.

Type of health care costs, expenses and reimbursements	n (%), Value
Outpatient costs	
Total outpatient expenses[a] (n=3000)	
¥0-¥4999 (US $0-$722)	727 (24.3)
¥5000-¥9999 (US $723-$867)	474 (15.8)
¥10,000-¥19,999 (US $1445-$2890)	547 (18.2)
over ¥20,000 (≥US $2890)	1252 (41.7)
Proportion of reimbursement (n=3000)	
Unknown	53 (1.8)
0%	1394 (46.5)
1%-49%	743 (24.7)
50%-99%	801 (26.7)
100%	9 (0.3)
Hospitalization costs	
Total hospitalization expenses (n=3000)	
0	857 (28.6)
¥1-¥9999 (US $0-$1445)	268 (8.9)
¥10,000-¥49,999 (US $1445-$7226)	955 (31.8)
over ¥50,000 (≥US $7226)	920 (30.7)
Proportion of reimbursement (n=2143)	
0%	140 (6.5)
1%-49%	832 (38.9)
50%-99%	1169 (54.5)
100%	2 (0.1)
Biological agent costs	
Frequency of use of biological agents (times) (n=3000)	
0	2072 (69.1)
1-9	874 (29.1)
Over 10	54 (1.8)
Proportion of reimbursement (n=1014)	
All at one’s own expense	626 (61.7)
Partial reimbursement	378 (37.3)
Complete reimbursement	10 (1.0)
Surgery costs	
Surgery costs (n=707)	
¥0-¥10,000 (US $0-$1445)	122 (17.3)
¥10,001-¥20,000 (US $1445-$2890)	129 (18.2)
¥20,001-¥49,999 (US $2890-$7226)	162 (22.9)
over ¥50,000 (≥US $7226)	294 (41.6)
Proportion of reimbursement (n=707)	
0%	40 (5.7)
1%-49%	295 (41.8)
50%-99%	370 (52.3)

Type of health care costs, expenses and reimbursements	n (%), Value
100%	2 (0.3)

[a] US $1=¥6.9197.

The costs of the surgeries were more than ¥50,000 (US $7226) for 41.6% (294/707) of the patients, but only 52.3% (370/707) of these patients reported receiving reimbursements of over 50% during the last 12 months. Additionally, 76.4% (2292/3000) of the patients had dietitian costs, including nutrition powder, nutrition solutions, or a specific carbohydrate diet during the last 12 months. Most of them (1300/3000, 43.3%) reported paying less than ¥20,000 (US $2890) during the last 12 months. However, 56.3% (1290/2292) of those who had dietitian costs paid these costs at their own expense without any reimbursement. Of note, for those who underwent special treatments such as fecal bacteria transplantation and stem cell transplantation, 30.2% (152/475) of them reported receiving no reimbursement.

Health Care Access

We assessed the health care access of patients with IBD by mainly focusing on the health and life insurance, primary care providers, emergency room visits, and hospitalization utilization. The majority of the respondents (2894/3000, 96.5%) had regular medical insurance. However, 75.3% (2259/3000) of the respondents had no private commercial insurance; the reasons for patients being without private commercial insurance included that they had been refused coverage by an insurance company after their diagnosis of IBD (1399/2576, 54.3%), they had not thought about commercial insurance (741/2576, 28.8%), or they had not decided to buy it yet (647/2576, 25.1%) (Figure 2A). Approximately half of the respondents (1282/3000, 42.7%) had no primary care provider (Table 1). Of those who were identified with a primary care provider, 73.5% (1262/1718) of them had a gastroenterologist, 68.8% (1182/1718) had an IBD expert, 11.8% (203/1718) had an anorectal surgeon, 7.2% (123/1718) had a gynecologist, 2.7% (46/1718) had an abdominal surgeon, 1.0% (18/1718) had community physicians, and 0.6% (10/1718) had a family practitioner. Of note, some patients with IBD had more than one primary care provider. Moreover, 84.4% (2532/3000) of the patients with IBD were without IBD surgeons, and of the 15.6% (468/3000) of the patients with IBD who had IBD surgeons, only 60.9% (285/468) of their surgeons were considered full mastery. As shown in Table 3, we further assessed the experience of emergency room visits and found that 68.0% (2040/3000) of the total surveyed patients had visited an emergency room each year after IBD diagnosis: 52.2% (1567/3000) reported going to the emergency room 1 to 4 times per year and 15.8% (473/3000) reported going to the emergency room more than 5 times. Moreover, 74.8% (2243/3000) of the patients reported having experienced hospitalization, with 38.1% (1144/3000) of them reporting 1-3 occurrences, 31.5% (946/3000) reporting 4-10 occurrences, and 5.1% (153/3000) reporting more than 10 occurrences.

Figure 2. Analysis of the feelings of patients with inflammatory bowel disease (IBD) with regard to the health care and the associated costs. A. Reasons for patients with IBD to not have commercial insurance (n=3000). B. Reasons identified by patients with IBD for delaying health care (n=1389). C. Patients' anxiety about medical expenses (n=3000). D. How patients with IBD save money and delay care (n=3000).

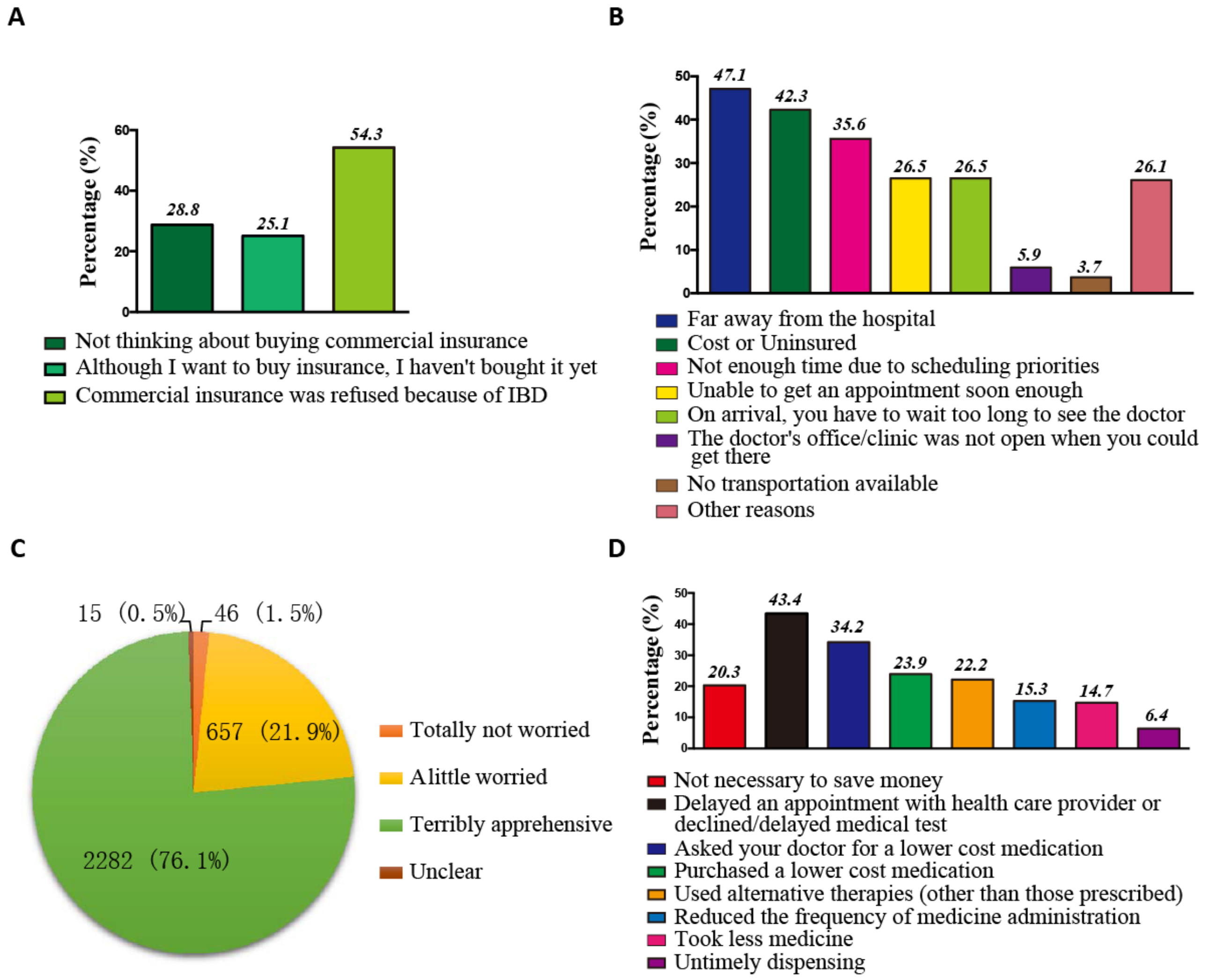


Table 3. Health care access of patients with inflammatory bowel disease (n=3000).

Health care access	n (%), Value
Average emergency visits per year after IBD[a]	
0	960 (32.0)
1-4 times	1567 (52.2)
Over 5 times	473 (15.8)
Hospitalization during last 12 months[b]	
0	757 (25.2)
1-3 times	1144 (38.1)
4-10 times	946 (31.5)
over 10 times	153 (5.1)
Average waiting time for hospitalization (days)	
1	443 (16.0)
2-7	1547 (55.9)
8-30	634 (22.9)
Over 30	143 (5.2)
Diagnosis and treatment of IBD in nearby medical institutions with nonspecialists in IBD treatment	
None	1411 (47.0)
Often	689 (23.0)
Once	220 (23.2)
Occasionally	203 (6.8)
Transportation facility for visiting the nearest IBD specialist	
Walking	22 (0.7)
Bicycle	51 (1.7)
Self-driving	420 (14.0)
Public transport	1462 (48.7)
Train	796 (26.5)
Airplane	17 (0.6)
Ship	1 (0.0)
Others	231 (7.7)

[a] IBD: inflammatory bowel disease.

[b] Injections of biological agents or their dispensing are all performed in the hospitals.

The optimal treatment of IBD requires specialized health care; however, patients might travel a long distance to obtain care for IBD, and a greater distance to the referral health care center was demonstrated to be associated with an increased risk of needing IBD-related surgery and outcomes [16]. As shown in Table 4, 42.2% (1266/3000) of the patients covered a distance of more than 100 km, and 73.2% (2197/3000) of the patients needed more than 1 hour to reach the closest IBD center. The transportation cost for 62.4% (1871/3000) of the patients to the nearest IBD center was more than ¥100 (US $14), and the accommodation cost for 75.4% (2262/3000) of the patients to the nearest IBD center was more than ¥100 (US $14). Consequently, 53.0% (1589/3000) of the patients with IBD reported going to nearby medical institutions without IBD specialists. As shown in Figure 2B, there were several reasons for patients to delay health care, including being far away from the hospital (654/1389, 47.1%), the cost or being uninsured (587/1389, 42.3%), not having enough time due to scheduling priorities (494/1389, 35.6%), being unable to get an appointment soon enough (368/1389, 26.5%), and having to wait too long to see the doctor after arrival (368/1389, 26.5%). When asked about how they felt about the financial cost (Figure 2C), 98.5% (2954/3000) of the patients with IBD were worried about the medical expenses. Nearly 79.7% (2392/3000) of the patients with IBD sought ways to save money for their health care. As shown in Figure 2D, 79.7% (2392/3000) of the IBD patients sought ways to save money for their health care. Almost 43.3% of the patients (1303/3000) delayed an appointment with a health care provider or declined/delayed medical tests, 34.2% of the patients (1026/3000) asked doctors for a lower cost medication, and 22.2% of the patients (667/3000) used alternative therapies (other than those prescribed), while 30% of the patients (900/3000) even reduced the administration frequency or dosage of their medicine or 6.4% of the patients (192/3000) engaged in the untimely dispensing of their medication.

Multifactor Analysis

We then focused on the influencing factors of emergency room visit times (Table 4). We found that patients with lower income (less than ¥5000 or US $723/month) were more likely to get into emergency rooms compared to patients with income more than ¥10,000 (US $1445) per month (odds ratio [OR] 1.947, 95% Cl 1.555-2.437; *P*<.001). Patients with lower transportation costs (¥0-¥99, US $0-$14) for visiting IBD specialists were less likely to visit emergency rooms compared to patients with higher transportation costs (>¥500, US $0-$72) (OR 0.622, 95% Cl 0.500-0.875; *P*=.004). Patients with ¥0-¥99 (US $0-$14) accommodation costs for visiting IBD specialists (OR 0.632, 95% Cl 0.494-0.810; *P*<.001) and ¥100-¥499 (US $14-$72) accommodation costs (OR 0.712, 95% Cl 0.584-0.866; *P*=.001) had less emergency room visit times than patients with >¥500 (US $72) accommodation costs (Table 4).

Table 4. Transportation costs and additional costs for medical treatments (n=3000).

Factor	n (%), Value	Emergency visit times		
		Odds ratio	95% Cl	*P* value
Distance to the nearest medical institution with IBD[a] specialists				.63
>100 km	1266 (42.2)	1	N/A[b]	N/A
10-99 km	1145 (38.2)	1.106	0.896-1.365	.35
<10 km	589 (19.6)	1.069	0.787-1.452	.67
Income of patients (per month)				*<.001*
>¥10,000 (US $1445)	434 (14.5)	1	N/A	N/A
¥5000-¥9999 (US $723-$1445)	1081 (36.0)	1.224	0.976-1.535	.08
¥0-¥4999 (US $0-$723)	1485 (49.5)	1.947	1.555-2.437	*<.001*
Time for visiting the nearest IBD specialist				.73
>6 hours	372 (12.4)	1	N/A	N/A
1-5 hours	1825 (60.8)	1.105	0.844-1.447	.47
<1 hour	803 (26.8)	1.060	0.745-1.508	.75
Total transportation costs for visiting IBD specialists[c]				*.02*
>¥500 (US $72)	634 (21.1)	1	N/A	N/A
¥100-¥499 (US $14-$72)	1234 (41.1)	0.816	0.650-1.024	.08
¥0-¥99 (US $0-$14)	1132 (37.7)	0.622	0.500-0.875	*.004*
Accommodation costs for visiting IBD specialists[c]				*<.001*
>¥500 (US $72)	791 (26.4)	1	N/A	N/A
¥100-¥499 (US $14-$72)	1467 (48.9)	0.712	0.584-0.866	*.001*
¥0-¥99 (US $0-$14)	742 (24.7)	0.632	0.494-0.810	*<.001*

[a] IBD: inflammatory bowel disease.

[b] N/A: not applicable.

[c] A multiple logistic regression model was used to obtain the *P* value. Values in italics are significant at *P*<.05.

Discussion

In conjunction with large at-risk populations, the absolute number of patients with IBD in newly industrialized countries has the potential to approximate that in the western world by 2025 [3]. IBD places an economic strain on health systems due to expensive pharmaceutical therapy, the risk of hospitalization and surgery, and long-term monitoring. Our study provides a detailed description of the financial burdens and health care access for patients with IBD in China. We found that heavy financial burdens due to health care and travelling long distances to receive health care were the main problems faced by Chinese patients with IBD at present.

Our conservative data showed the mean (SD) direct and indirect costs for per patient with IBD per year were US $11,668.68 ($7944.44) and US $74.90 ($253.60) by estimation. Recent studies reported that in the United States, the lifetime financial burden of Crohn disease or ulcerative colitis per person was estimated at US $622,000 to US $405,000. The total financial burden for Crohn disease and ulcerative colitis in 2016 was US $498 billion and US $377 billion, respectively [19]. In Canada, the economic cost for IBD was estimated at over US $3520 per person per year in 2018. Moreover, for rheumatoid arthritis in China, the mean (SD) direct and indirect costs were US $1917.21 ($2559.06) and US $492.88 ($1739.74) per patient year [20]. This did not include the intangible costs such as out of study or work. These findings demonstrate that IBD incurs a heavier financial burden than other chronic disease such as rheumatoid arthritis in China, with similar cost status reported in other western countries.

Our study showed that in over one-third of the surveyed patients, the outpatient and inpatient expenses exceeded ¥20,000 and ¥50,000 (US $2890 and US $7226, respectively) during the last 12 months. Only one-third of the inpatient costs were reimbursed at a rate of 60%, while the reimbursement ratio was even lower for the outpatient expenses for most patients. In addition, more than half of the patients who received biological agent treatment reported paying for the biological agents by themselves, and nearly half of the surgical costs had a reimbursement rate of less than 50%. However, the income of 49.5% (1485/3000) of the patients with IBD was less than ¥5000 (US $722) per month. Moreover, IBD affects the prime working years of patients with IBD and potentially has considerable effects on employment. IBD-related work losses have received great attention worldwide [21]. As shown above, 1907 (63.5%) of the 3000 patients with IBD in our survey reported not being able to work full-time.

Medical insurance policies in China vary widely between different provinces and are also different between urban and rural areas. In our study, 96.5% (2894/3000) of the patients had health insurance, 33.7% (1011/3000) had coverage under a new type of rural cooperative medical care system, and 5.2% (157/3000) had student health insurance. This is due to the government’s universal coverage of basic medical insurance for citizens. In China, social health insurance programs generally consist of 3 programs: the rural New Cooperative Medical Scheme launched in 2003, the Urban Residents Basic Medical Insurance program launched in 2007, and the Urban Employee Basic Medical Insurance program launched in 1998 [22]. Although insurance programs have rapidly expanded during the past decade, the benefit packages and deductibles vary between the programs. Rural populations have more restricted access to health care than urban residents do, with a larger financial burden, mainly due to a lower funding level for the New Cooperative Medical Scheme, for which the coverage of outpatient services is almost nonexistent [16,22-24]. However, 75.3% (2259/3000) of the surveyed patients did not have commercial insurance, with 54.3% (1399/2576) of the patients having been refused commercial insurance because of IBD. IBD is a chronic disease and cannot be cured, and the purchase of commercial private insurance for patients with IBD is generally rejected in China. Therefore, IBD-related costs have become a heavy financial burden for patients.

We also found that 76.1% (2282/3000) of the patients with IBD reported experiencing tremendous apprehension due to medical expenses and seeking ways to save money, such as delaying an appointment with a health care provider or declining/delaying medical tests, asking their doctors for a lower cost medication, purchasing a lower cost medication, using alternative therapies (other than those prescribed), reducing the frequency of medicine administration, taking less medicine, or engaging in untimely dispensing. However, all these are very harmful approaches to disease management, which in turn may lead to disease recurrence and further aggravating the economic pressures.

In our study, 52.2% (1567/3000) of the patients reported needing to go to the emergency department 1-4 times per year because of IBD. Both transportation costs and accommodation costs for visiting IBD specialists were positively associated with emergency room visit times. One possible reason may be that those patients with lower income level saved money by decreasing the follow-up times, which might lead to disease status becoming uncontrollable. The other reason could be that patients with lower income or without commercial private insurance cannot afford such expensive but effective biological agents, which resulted in a lower level of remission. This is consistent with that reported previously that travelling long distances to obtain medical treatment for IBD hinders regular care and adversely affects the outcomes—especially increasing the risk of needing IBD-related surgery [25].

Effective measures for solving the economic problems are reducing the expenses and improving the disease status for IBD. The reimbursement ratio for IBD should be increased especially for outpatient expenses, which may be difficult for this large population. Moreover, IBD experts are very scarce, especially in the city community. Approximately 42.7% (1282/3000) of the patients with IBD were without a primary care provider, and gastroenterologists and IBD experts were the most important primary care providers. Patients with IBD have difficulty obtaining medical services in China because community physicians account for 1% of the primary care providers and family practitioners account for 0.6% of the primary care providers [26]. As a result, 28.3% of the patients reported engaging in emergency dispensing. Therefore, telemedicine might be a good solution. In 2020, one web-based study showed that the use and need of telemedicine have been increasing especially since the outbreak of COVID-19 [27]. In the same year, we performed a web-based research [28] about patterns of care for patients with IBD in China during COVID-19; the results showed that a quarter of the patients sought care via telemedicine with their IBD physicians. One randomized controlled study conducted by de Jong et al [29] compared the cost-effectiveness of telemedicine and standard care for management of IBD; they found that telemedicine was safe and reduced outpatient visits and hospital admissions compared with standard care. Del Hoyo et al [30] suggested that compared with standard and telephone care, telemedicine with web-based programs decreased the direct and indirect costs for patients with IBD; other studies by Elkjaer [31] and de Jong [32] have shown similar results. Mao et al [33] mentioned that CCCF has organized a group of volunteer gastroenterologists that specialize in IBD to offer web-based consultancy to patients with IBD. Another study [34] reported that they sent educational and instructional alerts and messages to web-based IBD groups of patients via WeChat and mailed the drugs to the patients who lived far from the hospital. Therefore, telemedicine was shown to be not only cost-saving but also a medium to provide more accurate disease information for patients with IBD, which will also improve their medical adherence.

This is the first large-scale nationwide patient survey to assess the financial burdens and health care access of patients with IBD in China. Our results reflect the current status of Chinese patients with IBD and are vital for the national policy design for patients with IBD and the planning of CCCF projects in the future. Moreover, it is vital for IBD doctors and nurses to know more about the specific situation of patients with IBD in China and to understand more about the behaviors of patients with IBD. Our study plays a significant role in strengthening the disease management and improving the quality of life of patients with IBD in China. The imperfect and inherent self-report survey methodology, including its reliance on respondents to provide accurate and bias-free responses, was one limitation of our study. However, the aim of this survey was to assess patients' perception of access to health care and their financial burdens; thus, the self-report survey data are reasonable although they are imperfect. Moreover, patients with IBD who had limited access to the internet could not be included in our study, which was a potential source of selection bias in our study. While our study assessed patients with IBD in most regions of China, a large number of the included patients were mainly located in the provinces of Zhejiang, Anhui, and Jiangsu, wherein the economic status is higher than that in other areas; therefore, patients who could not access the internet may have worse economic status than our respondents. Thus, our study was just a mirror, reflecting part of the economic issues, but the reality is even worse than our estimation.

In conclusion, serious financial burdens and difficulties in health care access are the 2 major difficulties faced by patients with IBD in China. These problems need to be confronted by national health care systems, social force communities, and the patients themselves. Early purchase of private insurance and increasing the imbursement ratio of medical costs will cut the direct costs of IBD, and increasing use of telemedicine may be beneficial for decreasing both the direct and indirect costs for Chinese patients with IBD.

Acknowledgments

We thank the following professors for their great help in data collection: Dr. Xiaocang Cao (Department of Gastroenterology, Tianjin Medical University General Hospital), Dr. Min Chen (Department of Gastroenterology, Zhongnan Hospital of Wuhan University), Dr. Yihong Fan (Department of Gastroenterology, First Affiliated Hospital of Zhejiang University of Traditional Chinese Medicine), Dr. Yubei Gu (Department of Gastroenterology, Ruijin Hospital Affiliated to Medical College of Shanghai Jiaotong University), Dr. Hong Guo (Department of Gastroenterology, Second Affiliated Hospital of Military Medical University), Dr. Sun Jing (Department of Gastroenterology, Dr. Ruijin Hospital Affiliated to Medical College of Shanghai Jiaotong University), Dr. Yue Li (Department of Gastroenterology, Peking Union Medical University Hospital), Dr. Youyou Luo (Department of Gastroenterology, Children's Hospital Affiliated to Medical College of Zhejiang University), Dr. Yan Jia (Department of Gastroenterology, Former Army General Hospital of the Chinese People's Liberation Army), Dr. Chunhui Ouyang (Department of Gastroenterology, Second Xiangya Hospital of Central South University), Dr. Jun Shen (Department of Gastroenterology, Renji Hospital Affiliated to Medical College of Shanghai Jiaotong University, Shanghai Institute of Digestive Diseases, Shanghai Inflammatory Bowel Disease Research Center), Dr. Yuan Xiao (Department of Pediatrics, Ruijin Hospital Affiliated to Medical College of Shanghai Jiaotong University), and Dr. Min Zhi (Department of Gastroenterology, Sixth Hospital Affiliated to Sun Yat-sen University). We also thank Prof. Langting Lv (Chinese public health specialist, Health policy and management, Renmin University of China) for designing the questionnaire, critically reviewing, and providing constructive comments for this manuscript. This work was supported by research grants from National Natural Science Foundation of China, No. 81602516 and 81700455; Zhejiang Provincial National Science Foundation of China, No. LY20H160031, and Intestinal Barrier Research Foundation of Li Jieshou, No. LJS-201703.

Authors' Contributions

ZCP and CY designed the study. CY, ZCP, HSR, CHW, and YS collected and analyzed the data. YQ, XLY, and CY wrote the manuscript.

Conflicts of Interest

None declared.

References

1. Rubin DT, Feld LD, Goeppinger SR, Margolese J, Rosh J, Rubin M, et al. The Crohn's and Colitis Foundation of America Survey of Inflammatory Bowel Disease Patient Health Care Access. Inflamm Bowel Dis 2017 Feb;23(2):224-232. [doi: 10.1097/MIB.0000000000000994] [Medline: 27997434]
2. Chi KR. Epidemiology: Rising in the East. Nature 2016 Dec 21;540(7634):S100-S102. [doi: 10.1038/540S100a] [Medline: 28002397]
3. Kaplan GG. The global burden of IBD: from 2015 to 2025. Nat Rev Gastroenterol Hepatol 2015 Dec;12(12):720-727. [doi: 10.1038/nrgastro.2015.150] [Medline: 26323879]
4. Lönnfors S, Vermeire S, Greco M, Hommes D, Bell C, Avedano L. IBD and health-related quality of life -- discovering the true impact. J Crohns Colitis 2014 Oct;8(10):1281-1286. [doi: 10.1016/j.crohns.2014.03.005] [Medline: 24662394]

5. Stone CD. The economic burden of inflammatory bowel disease: clear problem, unclear solution. Dig Dis Sci 2012 Dec;57(12):3042-3044. [doi: 10.1007/s10620-012-2417-8] [Medline: 23086111]
6. Mehta F. Report: economic implications of inflammatory bowel disease and its management. Am J Manag Care 2016 Mar;22(3 Suppl):s51-s60 [FREE Full text] [Medline: 27269903]
7. Park KT, Colletti RB, Rubin DT, Sharma BK, Thompson A, Krueger A. Health Insurance Paid Costs and Drivers of Costs for Patients With Crohn's Disease in the United States. Am J Gastroenterol 2016 Jan;111(1):15-23. [doi: 10.1038/ajg.2015.207] [Medline: 26195179]
8. Kuenzig ME, Lee L, El-Matary W, Weizman AV, Benchimol EI, Kaplan GG, et al. The Impact of Inflammatory Bowel Disease in Canada 2018: Indirect Costs of IBD Care. J Can Assoc Gastroenterol 2019 Feb;2(Suppl 1):S34-S41 [FREE Full text] [doi: 10.1093/jcag/gwy050] [Medline: 31294383]
9. Burisch J, Jess T, Martinato M, Lakatos PL, ECCO -EpiCom. The burden of inflammatory bowel disease in Europe. J Crohns Colitis 2013 May;7(4):322-337. [doi: 10.1016/j.crohns.2013.01.010] [Medline: 23395397]
10. van der Valk ME, Mangen MJ, Leenders M, Dijkstra G, van Bodegraven AA, Fidder HH, COIN study groupthe Dutch Initiative on CrohnColitis. Healthcare costs of inflammatory bowel disease have shifted from hospitalisation and surgery towards anti-TNFα therapy: results from the COIN study. Gut 2014 Jan;63(1):72-79. [doi: 10.1136/gutjnl-2012-303376] [Medline: 23135759]
11. Kaplan GG, Ng SC. Understanding and Preventing the Global Increase of Inflammatory Bowel Disease. Gastroenterology 2017 Feb;152(2):313-321.e2. [doi: 10.1053/j.gastro.2016.10.020] [Medline: 27793607]
12. Ng SC, Zeng Z, Niewiadomski O, Tang W, Bell S, Kamm MA, Asia-Pacific Crohn'sColitis Epidemiology Study (ACCESS) Group. Early Course of Inflammatory Bowel Disease in a Population-Based Inception Cohort Study From 8 Countries in Asia and Australia. Gastroenterology 2016 Jan;150(1):86-95.e3; quiz e13. [doi: 10.1053/j.gastro.2015.09.005] [Medline: 26385074]
13. Ng SC, Kaplan GG, Tang W, Banerjee R, Adigopula B, Underwood FE, et al. Population Density and Risk of Inflammatory Bowel Disease: A Prospective Population-Based Study in 13 Countries or Regions in Asia-Pacific. Am J Gastroenterol 2019 Jan;114(1):107-115. [doi: 10.1038/s41395-018-0233-2] [Medline: 30177785]
14. Chen Y. Perspectives of IBD China: Is Crohn's and Colitis Foundation Model a Solution to Health Care Issues for the Country? Inflamm Bowel Dis 2018 Apr 23;24(5):925-929. [doi: 10.1093/ibd/izy056] [Medline: 29688396]
15. Luo X, Mao R, Chen B, Qiu Y, Zhang S, He Y, et al. Over-reaching beyond disease activity: the influence of anxiety and medical economic burden on health-related quality of life in patients with inflammatory bowel disease. Patient Prefer Adherence 2017;11:23-31 [FREE Full text] [doi: 10.2147/PPA.S118589] [Medline: 28053510]
16. Mak L, Ng SC, Wong IOL, Li MKK, Lo FH, Wong MTL, et al. Direct health-care cost utilization in Hong Kong inflammatory bowel disease patients in the initial 2 years following diagnosis. J Gastroenterol Hepatol 2018 Jan;33(1):141-149. [doi: 10.1111/jgh.13817] [Medline: 28475813]
17. Wenjuanxing: platform for survey design. URL: https://www.wjx.cn [accessed 2020-03-01]
18. Yu Q, Xu L, Li L, Zhi M, Gu Y, Wang X, et al. Internet and WeChat used by patients with Crohn's disease in China: a multi-center questionnaire survey. BMC Gastroenterol 2019 Jun 20;19(1):97 [FREE Full text] [doi: 10.1186/s12876-019-1011-3] [Medline: 31221086]
19. Lichtenstein GR, Shahabi A, Seabury SA, Lakdawalla DN, Espinosa OD, Green S, et al. Lifetime Economic Burden of Crohn's Disease and Ulcerative Colitis by Age at Diagnosis. Clin Gastroenterol Hepatol 2020 Apr;18(4):889-897.e10. [doi: 10.1016/j.cgh.2019.07.022] [Medline: 31326606]
20. Hu H, Luan L, Yang K, Li S. Burden of rheumatoid arthritis from a societal perspective: A prevalence-based study on cost of this illness for patients in China. Int J Rheum Dis 2018 Aug;21(8):1572-1580. [doi: 10.1111/1756-185X.13028] [Medline: 28211251]
21. Longobardi T, Jacobs P, Wu L, Bernstein CN. Work losses related to inflammatory bowel disease in Canada: results from a National Population Health Survey. Am J Gastroenterol 2003 Apr;98(4):844-849. [doi: 10.1111/j.1572-0241.2003.07378.x] [Medline: 12738466]
22. Meng Q, Fang H, Liu X, Yuan B, Xu J. Consolidating the social health insurance schemes in China: towards an equitable and efficient health system. Lancet 2015 Oct 10;386(10002):1484-1492. [doi: 10.1016/S0140-6736(15)00342-6] [Medline: 26466052]
23. You X, Kobayashi Y. The new cooperative medical scheme in China. Health Policy 2009 Jun;91(1):1-9. [doi: 10.1016/j.healthpol.2008.11.012] [Medline: 19121873]
24. Lei X, Lin W. The New Cooperative Medical Scheme in rural China: does more coverage mean more service and better health? Health Econ 2009 Jul;18 Suppl 2:S25-S46. [doi: 10.1002/hec.1501] [Medline: 19551752]
25. Borren NZ, Conway G, Tan W, Andrews E, Garber JJ, Yajnik V, et al. Distance to Specialist Care and Disease Outcomes in Inflammatory Bowel Disease. Inflamm Bowel Dis 2017 Jul;23(7):1234-1239 [FREE Full text] [doi: 10.1097/MIB.0000000000001133] [Medline: 28520589]
26. Shu Z, Wang Z, Chen R, Li M, Lou J, Huang X, et al. Allocation and development of the general practitioner workforce in China from 2012 to 2015: a literature review. The Lancet 2017 Dec;390:S91-S91 [FREE Full text] [doi: 10.1016/S0140-6736(17)33229-4]

27. Zhang Y, Qiu Y, He J, Tan J, Li X, Zhu L, et al. Impact of COVID-19 outbreak on the care of patients with inflammatory bowel disease: a comparison before and after the outbreak in South China. J Gastroenterol Hepatol 2020 Aug 01:1-10 [FREE Full text] [doi: 10.1111/jgh.15205] [Medline: 32738060]
28. Chen Y, Hu S, Wu H, Farraye FA, Bernstein CN, Zheng J, et al. Patterns of care for inflammatory bowel disease in China during the COVID-19 pandemic. Lancet Gastroenterol Hepatol 2020 Jul;5(7):632-634 [FREE Full text] [doi: 10.1016/S2468-1253(20)30131-X] [Medline: 32411921]
29. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. Lancet 2017 Sep 02;390(10098):959-968. [doi: 10.1016/S0140-6736(17)31327-2] [Medline: 28716313]
30. Del Hoyo J, Nos P, Bastida G, Faubel R, Muñoz D, Garrido-Marín A, et al. Telemonitoring of Crohn's Disease and Ulcerative Colitis (TECCU): Cost-Effectiveness Analysis. J Med Internet Res 2019 Sep 13;21(9):e15505 [FREE Full text] [doi: 10.2196/15505] [Medline: 31538948]
31. Elkjaer M, Shuhaibar M, Burisch J, Bailey Y, Scherfig H, Laugesen B, et al. E-health empowers patients with ulcerative colitis: a randomised controlled trial of the web-guided 'Constant-care' approach. Gut 2010 Dec;59(12):1652-1661. [doi: 10.1136/gut.2010.220160] [Medline: 21071584]
32. de Jong MJ, Boonen A, van der Meulen-de Jong AE, Romberg-Camps MJ, van Bodegraven AA, Mahmmod N, et al. Cost-effectiveness of Telemedicine-directed Specialized vs Standard Care for Patients With Inflammatory Bowel Diseases in a Randomized Trial. Clin Gastroenterol Hepatol 2020 Jul;18(8):1744-1752. [doi: 10.1016/j.cgh.2020.04.038] [Medline: 32335133]
33. Mao R, Liang J, Shen J, Ghosh S, Zhu L, Yang H, et al. Implications of COVID-19 for patients with pre-existing digestive diseases. Lancet Gastroenterol Hepatol 2020 May;5(5):425-427. [doi: 10.1016/s2468-1253(20)30076-5]
34. An P, Ji M, Ren H, Su J, Ding NS, Kang J, et al. Prevention of COVID-19 in patients with inflammatory bowel disease in Wuhan, China. Lancet Gastroenterol Hepatol 2020 Jun;5(6):525-527 [FREE Full text] [doi: 10.1016/S2468-1253(20)30121-7] [Medline: 32311321]

Abbreviations

CCCF: China Crohn’s & Colitis Foundation
IBD: inflammatory bowel disease
OR: odds ratio

Edited by G Eysenbach; submitted 24.05.20; peer-reviewed by J del Hoyo Francisco, Q Yang; comments to author 22.06.20; revised version received 30.08.20; accepted 26.10.20; published 05.01.21.

Please cite as:
Yu Q, Zhu C, Feng S, Xu L, Hu S, Chen H, Chen H, Yao S, Wang X, Chen Y
Economic Burden and Health Care Access for Patients With Inflammatory Bowel Diseases in China: Web-Based Survey Study
J Med Internet Res 2021;23(1):e20629
URL: https://www.jmir.org/2021/1/e20629
doi:10.2196/20629
PMID:33399540

Original Paper

Retracted: "Assessing the Dissemination of COVID-19 Articles Across Social Media With Altmetric and PlumX Metrics: Correlational Study"

Haley N Tornberg[1*], BS; Carine Moezinia[1*], MD; Chapman Wei[2], BS; Simone A Bernstein[3], MD; Chaplin Wei[4], BS; Refka Al-Beyati[5], MD; Theodore Quan[2], BS; David J Diemert[2], MD, FRCP

[1]Department of Medicine, Hospital for Special Surgery, New York, NY, United States
[2]Department of Medicine, The George Washington School of Medicine and Health Sciences, Washington, DC, United States
[3]Department of Psychiatry, Washington University School of Medicine and Barnes-Jewish Hospital, St. Louis, MO, United States
[4]Department of Medicine, American University of Antigua, Coolidge, Antigua and Barbuda
[5]Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, CA, United States
[*]these authors contributed equally

Corresponding Author:
Refka Al-Beyati, MD
Department of Medicine
David Geffen School of Medicine at University of California
10833 Le Conte Ave
Los Angeles, CA
United States
Phone: 1 3133102931
Email: Ralbeyat@gmail.com

Related Article:

Do not cite this article - this article has been retracted. Retraction statement: https://www.jmir.org/2022/8/e41544/

Abstract

Background: The use of social media assists in the distribution of COVID-19 information to the general public and health professionals. Alternative-level metrics (ie, altmetrics) and PlumX metrics are new bibliometrics that can assess how many times a scientific article has been shared and how much a scientific article has spread within social media platforms.

Objective: Our objective was to characterize and compare the traditional bibliometrics (ie, citation count and impact factors) and new bibliometrics (ie, Altmetric Attention Score [AAS] and PlumX score) of the top 100 COVID-19 articles with the highest AASs.

Methods: The top 100 articles with highest AASs were identified with Altmetric Explorer in May 2020. The AASs, journal names, and the number of mentions in various social media databases of each article were collected. Citation counts and PlumX Field-Weighted Citation Impact scores were collected from the Scopus database. Additionally, AASs, PlumX scores, and citation counts were log-transformed and adjusted by +1 for linear regression, and Spearman correlation coefficients were used to determine correlations.

Results: The median AAS, PlumX score, and citation count were 4922.50, 37.92, and 24.00, respectively. The New England Journal of Medicine published the most articles (18/100, 18%). The highest number of mentions (985,429/1,022,975, 96.3%) were found on Twitter, making it the most frequently used social media platform. A positive correlation was observed between AAS and citation count (r^2=0.0973; P=.002), and between PlumX score and citation count (r^2=0.8911; P<.001).

Conclusions: Our study demonstrated that citation count weakly correlated with AASs and strongly correlated with PlumX scores, with regard to COVID-19 articles at this point in time. Altmetric and PlumX metrics should be used to complement traditional citation counts when assessing the dissemination and impact of a COVID-19 article.

(J Med Internet Res 2021;23(1):e21408) doi:10.2196/21408

KEYWORDS

Altmetric; PlumX; social media; impact factor; COVID-19; information; dissemination; citation

Introduction

The SARS-CoV-2 virus is the pathogen responsible for the latest global pandemic that has exhausted the global economy and health care system to a degree that has not been seen since the 1918 influenza outbreak. The virus originated from Wuhan, China in December 2019, and as of May 1, 2020, more than 230,000 COVID-19 fatalities have been reported worldwide, which is higher than the number of fatalities reported for both severe acute respiratory syndrome and Middle East respiratory syndrome combined [1,2]. Due to the rapid spread of the virus and the massive number of casualties, there has been a rapid rate of research dissemination across medical journals and social media platforms to provide real-time guidance for understanding the epidemiology, disease characteristics, and clinical management of, and future treatment development for, COVID-19 to all stakeholders who are invested in managing the COVID-19 pandemic [3,4].

In contrast to medical journals, social media can serve as a useful platform for informing the wider general public (ie, both medical professionals and laypeople alike) and disseminating crucial and novel information during this evolving crisis [5,6]. To capture the level of an article's dissemination across social media (ie, an article's "online attention"), many metric tools, such as Altmetric and PlumX, have been created [7]. As opposed to the traditional metrics of article dissemination, such as article citation count, metrics that describe article dissemination across social media are known as alternative metrics or "altmetrics." The Altmetric Attention Score (AAS), which was developed by Altmetric, is a weighted score of the amount of "online attention" a research article has received across social media platforms. This score solely refers to the number of citations, linkouts, and abstract views, and such social media platforms include Twitter, Facebook, Google+, Wikipedia, blogs, and many others [8]. Similarly, PlumX has developed the PlumX Field-Weighted Citation Impact score (ie, PlumX score), which is also a weighted metric score of the level of article dissemination across similar social media platforms, based on the number of citations, linkouts, and abstract views [9].

Given that AASs and PlumX scores use relevant information from social media platforms, including Twitter and Facebook, these scores can be potentially useful adjunctive metric tools for holistically evaluating an article's impact or effect on a field of research, instead of just evaluating scholarly impact [10]. These impacts or effects include information uptake, information engagement, and the relevance of results. Citation counts and impact factors reflect the number of citations in other articles or journals, whereas altmetrics reflect the instantaneous attention that an article garners among news outlets, blogs, Twitter, Facebook, and other media platforms [4]. Additionally, since paper journals are moving toward internet-based platforms, the development of these new internet-based technologies provides researchers with a new approach to assessing the effect of research [4]. In terms of biomedical research, the use of Twitter to disseminate article information has increased dramatically over the years. This showcases the degree of social media use in the medical community [11]. Previous studies have assessed the utility of new bibliometrics (ie, AASs and PlumX scores) as complements to citation count in various medical fields, but these studies have reported variable results [4,9,10,12,13]. Furthermore, the utility of altmetrics in COVID-19 research has not yet been evaluated.

Given the lack of research, the massive influx of COVID-19 publications since early 2020, and the fact that altmetrics allow for the rapid assessment of an article's level of dissemination upon publication, altmetrics have the potential to be used complementarily with traditional bibliometrics (eg, article citation), which typically take years to accumulate [8]. The purpose of this study was to determine the utility of adjunctively using AAS and PlumX altmetrics as complements to traditional bibliometrics in the assessment of the 100 most "trending" COVID-19 articles across social media, as determined by Altmetric. Given that both the scientific community and the general public frequently check for new information in journals and social media platforms to gain a better understanding of how to prevent and manage COVID-19 [14], we hypothesized that there might be a significant correlation between AAS and PlumX metrics and citation count. Such a correlation would suggest an alignment between the interests of academic scholars and the general public. The primary objective of this study was to investigate the correlations between both Altmetric and PlumX scores and citation count.

Methods

Altmetric Explorer was used to identify COVID-19 publications from December 2019 to May 2020, by using the PubMed search terms "COVID-19," "SARS CoV-2," and "coronavirus." Articles that were retracted by May 2020 were excluded from analysis. The top 100 articles with the highest AASs were selected from the list of included articles. The number of mentions from the following Altmetric data components were extracted and examined: news mentions, blog mentions, policy mentions, Twitter mentions, Facebook mentions, Wikipedia mentions, Reddit mentions, Mendeley readers, and the number of Dimensions citations [9]. In addition to the extraction of Altmetric data components, we also collected data on the impact factors of the journals that the included articles were published in, article type, and article citation count [8,10]. For traditional citation analysis, article citation counts were found by using the Scopus database. To assess whether the results from the Altmetric analysis could be generalized across multiple social media metric tools, PlumX scores were collected from the Scopus database. AASs, PlumX scores, and citation counts were log-transformed and adjusted by +1 for linear regression, and Spearman correlation coefficients were used to determine correlations. Statistical significance was defined as $P<.05$.

Results

The majority of articles were published in biomedical journals (Table 1, Textbox 1). Compared to the other journals, the New England Journal of Medicine published the most articles (18/100, 18%). Of the 100 articles, 42 (42%) were original investigations. The article with the highest AAS (AAS=33,828) was a biomechanistic basic science letter that delineated the features of the SARS-CoV-2 genome and provided evidence that SARS-CoV-2 was not constructed in a laboratory (AAS=33,828; citation count=30) [15]. The article with the highest PlumX score was a prospective cohort study that described the clinical characteristics of patients with COVID-19 [16]. The median AAS, PlumX score, and citation count were 4,922.50, 37.92, and 24.00, respectively. All articles were classified as the top 5% in terms of scientific output [10]. This means that in terms of AAS, these articles ranked in the top 5% when compared to more than 15.5 million research publications. In other words, these articles garnered the most attention. A total of 1,022,975 mentions in social media platforms were assessed. The highest number of mentions for the selected articles (985,429/1,022,975, 96.3%) were found on Twitter. Additionally, 99 articles were open access. In total, 9283 Mendeley mentions and 18,011 Dimensions citations were present. A weak positive correlation was observed between AAS and citation count (r^2=0.0973; P=.002) (Figure 1). However, a strong positive correlation was observed between PlumX score and citation count (r^2=0.8911; P<.001). Stronger positive correlations were observed between the number of Mendeley readers (r^2=0.958; P<.001) and citation count, and between the number of Dimensions citations (r^2=0.984; P<.001) and citation count.

Table 1. Characteristics and components of the top 100 COVID-19 articles with the highest Altmetric Attention Scores.

Characteristic	Value
Altmetric Attention Score, median (range)	4922.50 (2841-33828)
PlumX score, median (range)[a]	37.92 (0-1862.23)
Journal impact factor, median (range)	41.06 (1.29-70.67)
Traditional citation count, median (range)[b]	24.00 (0-1096)
Number of news mentions, total (range)	32509 (1-2021)
Number of blog mentions, total (range)	2630 (0-131)
Number of policy mentions, total (range)	154 (0-21)
Number of Twitter mentions, total (range)	985429 (1381-84022)
Number of Facebook mentions, total (range)	1138 (0-58)
Number of Wikipedia mentions, total (range)	177 (0-11)
Number of Reddit mentions, total (range)	938 (0-40)
Number of Mendeley Readers, total (range)	9283 (0-2581)
Number of Dimensions citations, total (range)	18011 (0-2233)
Article type, n (%)	
Original investigation	44 (44)
Correspondence	31 (31)
Editorial	14 (14)
Review	4 (4)
Viewpoint	6 (6)
Open access	100 (100)
Study design of original investigations (n=44) , n (%)	
Clinical trials	4 (9)
Prospective/retrospective cohort studies	21 (48)
Cross sectional	1 (2)
Case series	7 (16)
Basic science in vitro/in vivo studies	8 (18)
Model validation studies	3 (7)

[a] In total, 16 articles had a PlumX score of 0, because they did not have a score on Scopus at the time of this study. Therefore, these articles were not part of the analysis.

[b] In total, 3 articles did not have a citation count on Scopus or PubMed Central at the time of this study. Therefore, these articles were not part of the analysis.

Textbox 1. Journals categorized based on journal impact factor. The number of included articles published in each journal are reported.

Journals with an impact factor of >40

- New England Journal of Medicine (articles: n=18)
- The Lancet (articles: n=14)
- Journal of the American Medical Association (articles: n=12)
- Nature (articles: n=2)
- Science (articles: n=6)

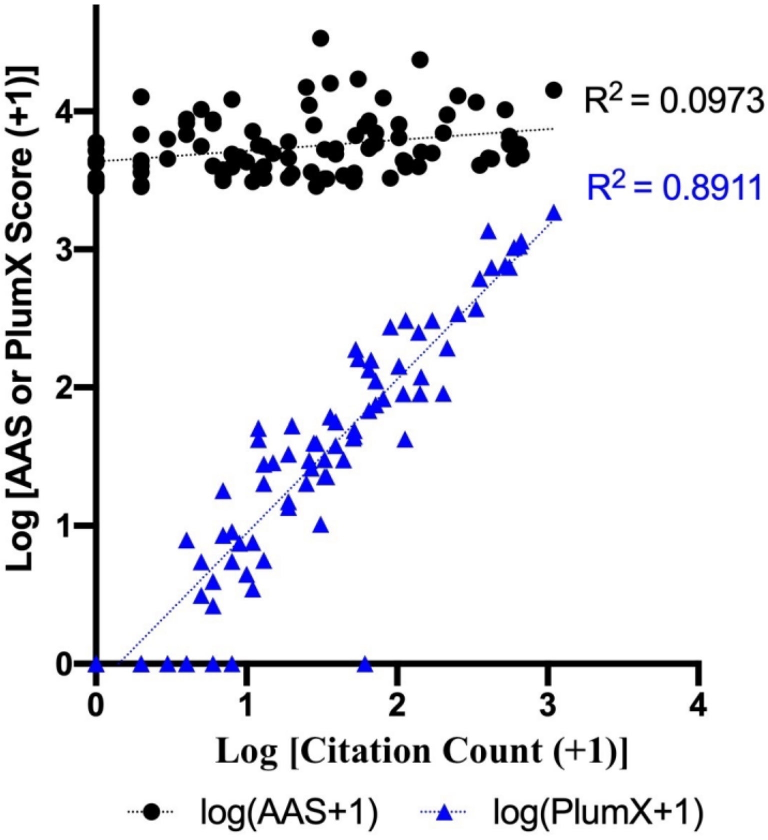
Journals with an impact factor of 20-40

- Cell (articles: n=1)
- Nature Biotechnology (articles: n=1)
- Nature Medicine (articles: n=5)
- British Medical Journal (articles: n=2)
- Lancet Infectious Disease (articles: n=3)
- The Lancet Respiratory Medicine (articles: n=2)

Journals with an impact factor of <20

- Annals of Internal Medicine (articles: n=2)
- Cell Research (articles: n=1)
- Morbidity and Mortality Weekly Report (articles: n=8)
- ACS (American Chemical Society) Nano (articles: n=1)
- Proceedings of the National Academy of Sciences of the United States of America (articles: n=2)
- Emerging Infectious Diseases (articles: n=4)
- Biomedicine (articles: n=1)
- Eurosurveillance (articles: n=1)
- Pediatrics (articles: n=1)
- Circulation: Arrhythmia and Electrophysiology (articles: n=1)
- International Journal of Antimicrobial Agents (articles: n=1)
- Cell Discovery (articles: n=1)
- Engineering (articles: n=1)
- Journal of Travel Medicine (articles: n=1)
- Antiviral Research (articles: n=1)
- Journal of Hospital Infection (articles: n=1)
- Journal of Medical Virology (articles: n=2)
- BioScience Trends (articles: n=1)
- The Lancet Child & Adolescent Health (articles: n=1)
- Médecine et Maladies Infectieuses (articles: n=1)
- JAMA (Journal of the American Medical Association) Network Open (articles: n=1)

Figure 1. Correlations between citation count and AAS and between citation count and PlumX score. AAS: Altmetric Attention Score.



Discussion

In the past several months since COVID-19 has spread globally, many articles pertaining to the pandemic have been disseminated. Traditionally, high-quality and highly regarded articles can be ascertained via citation count; however, this method is not always possible in a rapidly evolving pandemic, such as the COVID-19 pandemic. Our study assessed the utility of adjunctively using AASs and PlumX scores with citation count in the evaluation of the top 100 "trending" COVID-19 articles. According to the results of our study, AAS weakly correlated with citation count (r^2=0.0973; *P*=.002), whereas PlumX score strongly correlated with citation count (r^2=0.8911; *P*<.001). This suggests that there is an alignment in information dissemination between both peers within the scientific community and those in the general public.

Our results are consistent with the findings of some of the previous literature and inconsistent with those of other literature. Previous studies have shown that PlumX and Altmetric scores positively correlate with traditional citation count [12,17-21]. In more current literature, citation count has been shown to weakly positively correlate with PlumX score and AAS. However, PlumX and AAS measures have not consistently positively correlated with impact factor. It has also been shown that compared to all other alternative metric platforms, PlumX is able to capture the widest range of data regarding the dissemination and popularity of a scientific article [7]. Interestingly, current literature has noted that journals with a high Twitter presence also have high PlumX scores and AASs [22]. Our data also reinforces this finding. In our analysis, we observed that Twitter was the most frequently used social media platform, and that the highest number of mentions for the selected articles (985,429/1,022,975, 96.3%) were found on Twitter. These results are also in line with those of previous studies [23]. All of these studies have stated that PlumX score and AAS are most effective when they are used complementarily with traditional bibliometric measures, and that PlumX score and AAS should not necessarily be used on their own to assess an article's quality [4,7-10].

Although AASs can be used to estimate how widely an article has been disseminated, its correlation with citation count was lower than the correlation between PlumX scores and citation count in our study, making AASs less useful than PlumX scores. There are likely multiple reasons for this. For instance, there are several challenges with regard to the utility of AASs, such as the heterogeneity of the platforms that Altmetric uses to calculate the score, the dynamic nature of AASs, and the fact that many of the platforms we assessed (eg, Twitter, Facebook, etc) were not focused on academia [4]. These challenges allow for the possibility of article misinformation and the review of an article that may be retracted a few days later, which may why explain the weak correlation between AAS and citation count. Based on the platforms that Altmetric uses to calculate its weighted score and those that the scientific community finds interesting, it was not surprising to observe the strong correlations between citation count and Mendeley citations, and between citation count and Dimensions citations. Mendeley and Dimensions are 2 well-known platforms that are used by scientific researchers to source, organize, and cite preexisting literature. These platforms attract a more professional group of users who typically work in the biomedical science field compared to other platforms (ie, Twitter or Facebook), which attract a broader demographic group of the general public. Furthermore, although PlumX scores are also dynamic, these scores account for citation count in their weights, which make PlumX scores longer to report and more stable than other bibliometrics [9].

It is important to note that original articles, particularly observational cohort studies, were disseminated more quickly in social media platforms than in scientific journals. Furthermore, the top-scoring biomechanistic basic science article [15] was also rapidly disseminated in social media platforms. This is most likely because this article addressed the possibility of SARS-CoV-2 being engineered in a laboratory [15]. Additionally, all the articles were open access and freely accessible to the general public. This is likely because there is an alignment between scientists and the general public concerning the need to disseminate new COVID-19–related findings as quickly as possible, given the enormous burden of the ongoing COVID-19 pandemic. Thanks to the removal of barriers to access, the reading of newly published articles and the dissemination of articles to peers has considerably increased.

There are several limitations to our study. For instance, Altmetric and PlumX scores do not necessarily reflect the scientific quality of an article [12,24]. Just because an article generated more attention among the general public and academic community than other articles, it does not mean that more attention correlates with better study designs, results, or evidence quality. Additionally, AASs and PlumX scores are more dynamic than citation count, making the precision, consistency, and reproducibility of altmetric analyses difficult to achieve. However, due to the high volume of social media attention that COVID-19 articles have received, the reproducibility of our data is much better than the reproducibility of data from other medical fields, as the power of our results was much stronger. Despite these limitations, our study shows that altmetrics can be used to complement citation analysis for COVID-19 articles. We highly recommend that medical providers, the scientific community, and the general public use AASs when initially searching for the most pertinent articles that the general public is interested in, given that AASs are immediately calculated. The initial search should be followed by the assessment of the utility of the articles, with respect to PlumX scores or citation count. Given the high correlation between PlumX scores and citation count, we believe that this method will result in great confidence when reviewing a high-quality COVID-19 article and a low risk of article misinformation. We believe that our findings will be helpful in the current COVID-19 pandemic and in future pandemics, as traditional metrics, such as citation count, are not readily available when assessing an article's quality.

Conflicts of Interest

None declared.

References

1. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Center for Systems Science and Engineering at Johns Hopkins University. URL: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2020-05-01]
2. Mahase E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ 2020 Feb 18;368:m641. [doi: 10.1136/bmj.m641] [Medline: 32071063]
3. Song P, Karako T. COVID-19: Real-time dissemination of scientific information to fight a public health emergency of international concern. Biosci Trends 2020 Mar 16;14(1):1-2. [doi: 10.5582/bst.2020.01056] [Medline: 32092748]
4. Patthi B, Prasad M, Gupta R, Singla A, Kumar JK, Dhama K, et al. Altmetrics - A Collated Adjunct Beyond Citations for Scholarly Impact: A Systematic Review. J Clin Diagn Res 2017 Jun;11(6):ZE16-ZE20 [FREE Full text] [doi: 10.7860/JCDR/2017/26153.10078] [Medline: 28764311]
5. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
6. Campbell L, Evans Y, Pumper M, Moreno MA. Social media use by physicians: a qualitative study of the new frontier of medicine. BMC Med Inform Decis Mak 2016 Jul 15;16:91 [FREE Full text] [doi: 10.1186/s12911-016-0327-y] [Medline: 27418201]
7. Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz J. Research data explored: an extended analysis of citations and altmetrics. Scientometrics 2016;107:723-744 [FREE Full text] [doi: 10.1007/s11192-016-1887-4] [Medline: 27122647]
8. Elmore SA. The Altmetric Attention Score: What Does It Mean and Why Should I Care? Toxicol Pathol 2018 Apr;46(3):252-255 [FREE Full text] [doi: 10.1177/0192623318758294] [Medline: 29448902]
9. Ortega JL. Disciplinary differences of the impact of altmetric. FEMS Microbiol Lett 2018 Apr 01;365(7):1-6. [doi: 10.1093/femsle/fny049] [Medline: 29518193]
10. Patel RB, Vaduganathan M, Bhatt DL, Bonow RO. Characterizing High-Performing Articles by Altmetric Score in Major Cardiovascular Journals. JAMA Cardiol 2018 Dec 01;3(12):1249-1251 [FREE Full text] [doi: 10.1001/jamacardio.2018.3823] [Medline: 30476952]
11. Haustein S, Costas R, Larivière V. Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. PLoS One 2015;10(3):e0120495. [doi: 10.1371/journal.pone.0120495] [Medline: 25780916]

12. Jia JL, Nguyen B, Mills DE, Polin DJ, Sarin KY. Comparing online engagement and academic impact of dermatology research: An Altmetric Attention Score and PlumX Metrics analysis. J Am Acad Dermatol 2020 Aug;83(2):648-650. [doi: 10.1016/j.jaad.2019.12.003] [Medline: 31836559]
13. Parrish JM, Jenkins NW, Brundage TS, Hrynewycz NM, Singh K. The Top 100 Spine Surgery Articles on Social Media: An Altmetric Study. Spine (Phila Pa 1976) 2020 Sep 01;45(17):1229-1238. [doi: 10.1097/BRS.0000000000003508] [Medline: 32355144]
14. Ong EK, Lim CH, Wong AJ. The Role of Social Media During the COVID-19 Pandemic. Ann Acad Med Singap 2020 Jun;49(6):408-410 [FREE Full text] [Medline: 32712642]
15. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
16. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
17. Maggio LA, Leroux TC, Meyer HS, Artino Jr AR. #MedEd: exploring the relationship between altmetrics and traditional measures of dissemination in health professions education. Perspect Med Educ 2018 Aug;7(4):239-247 [FREE Full text] [doi: 10.1007/s40037-018-0438-5] [Medline: 29949099]
18. Chang J, Desai N, Gosain A. Correlation Between Altmetric Score and Citations in Pediatric Surgery Core Journals. J Surg Res 2019 Nov;243:52-58. [doi: 10.1016/j.jss.2019.05.010] [Medline: 31154133]
19. Nocera AP, Boyd CJ, Boudreau H, Hakim O, Rais-Bahrami S. Examining the Correlation Between Altmetric Score and Citations in the Urology Literature. Urology 2019 Dec;134:45-50. [doi: 10.1016/j.urology.2019.09.014] [Medline: 31560915]
20. Amath A, Ambacher K, Leddy JJ, Wood TJ, Ramnanan CJ. Comparing alternative and traditional dissemination metrics in medical education. Med Educ 2017 Sep;51(9):935-941. [doi: 10.1111/medu.13359] [Medline: 28719136]
21. Ruan QZ, Chen AD, Cohen JB, Singhal D, Lin SJ, Lee BT. Alternative Metrics of Scholarly Output: The Relationship among Altmetric Score, Mendeley Reader Score, Citations, and Downloads in Plastic and Reconstructive Surgery. Plast Reconstr Surg 2018 Mar;141(3):801-809. [doi: 10.1097/PRS.0000000000004128] [Medline: 29481413]
22. Dardas LA, Woodward A, Scott J, Xu H, Sawair FA. Measuring the social impact of nursing research: An insight into altmetrics. J Adv Nurs 2019 Jul;75(7):1394-1405. [doi: 10.1111/jan.13921] [Medline: 30507052]
23. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
24. Bornmann L, Haunschild R. Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data. PLoS One 2018;13(5):e0197133. [doi: 10.1371/journal.pone.0197133] [Medline: 29791468]

Abbreviations

AAS: Altmetric Attention Score

Edited by G Eysenbach; submitted 13.06.20; peer-reviewed by A Gu, K Reuter; comments to author 13.07.20; revised version received 16.08.20; accepted 12.12.20; published 14.01.21.

Please cite as:
Tornberg HN, Moezinia C, Wei C, Bernstein SA, Wei C, Al-Beyati R, Quan T, Diemert DJ
Retracted: "Assessing the Dissemination of COVID-19 Articles Across Social Media With Altmetric and PlumX Metrics: Correlational Study"
J Med Internet Res 2021;23(1):e21408
URL: http://www.jmir.org/2021/1/e21408/
doi:10.2196/21408
PMID:33406049

Original Paper

The HealthChain Blockchain for Electronic Health Records: Development Study

Yonggang Xiao[1], PhD; Bin Xu[1], PhD; Wenhao Jiang[2], PhD; Yunjun Wu[3], PhD

[1]School of Computer Science and Technology, Hubei University of Science and Technology, Xianning, China

[2]Chongqing Aerospace Polytechnic, Chongqing, China

[3]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China

Corresponding Author:
Bin Xu, PhD
School of Computer Science and Technology
Hubei University of Science and Technology
Xianning Road
Xianning, 437000
China
Phone: 86 13986628202
Email: jwcxb@hbust.edu.cn

Abstract

Background: Health care professionals are required to maintain accurate health records of patients. Furthermore, these records should be shared across different health care organizations for professionals to have a complete review of medical history and avoid missing important information. Nowadays, health care providers use electronic health records (EHRs) as a key to the implementation of these goals and delivery of quality care. However, there are technical and legal hurdles that prevent the adoption of these systems, such as concerns about performance and privacy issues.

Objective: This study aimed to build and evaluate an experimental blockchain for EHRs, named HealthChain, which overcomes the disadvantages of traditional EHR systems.

Methods: HealthChain is built based on consortium blockchain technology. Specifically, three organizations, namely hospitals, insurance providers, and governmental agencies, form a consortium that operates under a governance model, which enforces the business logic agreed by all participants. Every peer node hosts an instance of the distributed ledger consisting of EHRs and an instance of chaincode regulating the permissions of participants. Designated orderers establish consensus on the order of EHRs and then disseminate blocks to peers.

Results: HealthChain achieves functional and nonfunctional requirements. It can store EHRs in a distributed ledger and share them among different participants. Moreover, it demonstrates superior features, such as privacy preservation, security, and high throughput. These are the main reasons why HealthChain is proposed.

Conclusions: Consortium blockchain technology can help to build new EHR systems and solve the problems that prevent the adoption of traditional systems.

(J Med Internet Res 2021;23(1):e13556) doi:10.2196/13556

KEYWORDS

electronic health record; distributed ledger; privacy preservation; proof of authority; chaincode application programming interface

Introduction

It has long been believed that electronic health records (EHRs) should be maintained across time and space, and could be accessed at any time and any place within the law [1]. In the first stage of digitization, we store a patient's medical history within the jurisdiction of a health care provider irrespective whether electronic medical records (EMRs) are on a local server [2] or in the cloud [3]. Such EMR systems have no essential difference with old-fashioned paper-based ones, since information technology just takes the management of medical records from paper folders to hard drives. EMRs do not travel out of a practice [4], and they make their way to other practices by faxes or signed documents, which is time consuming.

In the second stage, authorized doctors and staff create, manage, and consult EHRs across more than one health care organization, allowing interoperability between disparate EHR systems [5]. That is, EHRs possess the ability to share medical information among health care providers and follow a patient's information across multiple health care organizations [6]. In the United States, EHR exchange involves a common platform, the Nationwide Health Information Network, which is a set of standards, services, and policies that enable secure health information exchange over the internet [7].

However, there are technical and legal hurdles that prevent the adoption of these systems. First, these systems perform poorly in terms of data availability, data integrity, and retrieval rate when EHRs are stored under a distributed or institution-centric model [8]. Second, people always worry about the issues of privacy and data breaches [9] when EHRs are beyond their control, even if health care providers and governmental agencies claim that these systems are Health Insurance Portability and Accountability Act (HIPAA) compliant [10]. After all, 11,581,616,452 records have been breached since 2005, and this has been reported through either government agencies or verifiable media sources [11]. Third, patient-reported data do not always get recorded in a patient's EHRs since doctor-patient communication is not always possible, which impacts the quality of care [12]. Therefore, a kind of patient-reporting mechanism is needed for precision medicine [13].

In this paper, HealthChain is proposed to address the above-mentioned issues. It is a blockchain [14-16] for EHRs, that is, a growing list of blocks that consist of records and are linked using a cryptographic hash [17]. The blockchain has several advantages. First, the blockchain is a distributed peer-to-peer database where data availability, data integrity, and response time are guaranteed [18,19]. Blockchains can facilitate Internet of Things security in eHealth [20]. Second, the blockchain operates under a governance model, which enforces the business logic agreed by all participants. Therefore, we can exploit a smart contract or chaincode to regulate the access control policy [21-23] and achieve HIPAA compliance. Third, the blockchain is managed collectively by its stakeholders, some of whom have the right to record data in the block that cannot be altered retroactively [24]. In our design, even patients can report personal health records (PHRs) [5] on the ledger. From our perspective, applying blockchain technology to EHR systems denotes the advent of the third stage of digitization.

Specifically, the proposed HealthChain is different from other EHR blockchains owing to the following features. First, HealthChain is a consortium blockchain [25]. Multiple organizations, namely hospitals, insurance providers, and governmental agencies, come together to form the consortium. The business logic is determined by the governance model that is agreed by the consortium at the beginning, rather than the trustless model of other medical blockchains [26-28]. Second, HealthChain performs well in the following aspects: data availability, data integrity, and retrieval success rate, that is, HealthChain is always online even if a few servers crash. Genuine EHRs are stored since they are signed by valid stakeholders. We can successfully access the ledger anytime because of load balancing. Third, HealthChain uses proof of authority (PoA) as its consensus protocol. Designated, authenticated, and trustworthy orderers are responsible for generating valid blocks, that is, as long as the blocks are signed by one of these orderers, they are accepted by all participants. PoA is different from other consensus protocols, such as proof of work [21,29,30] and Practical Byzantine Fault Tolerance (PBFT) [31-33]. Fourth, different users possess different chaincode application programming interfaces (APIs) in HealthChain, which is specified by the governance model. Therefore, we can define the way that users interact with the ledger and achieve the access control policy. Fifth, HealthChain performs well in the experimental environment. This paper uses the following four metrics to evaluate performance: read latency, read throughput, transaction latency, and transaction throughput [34].

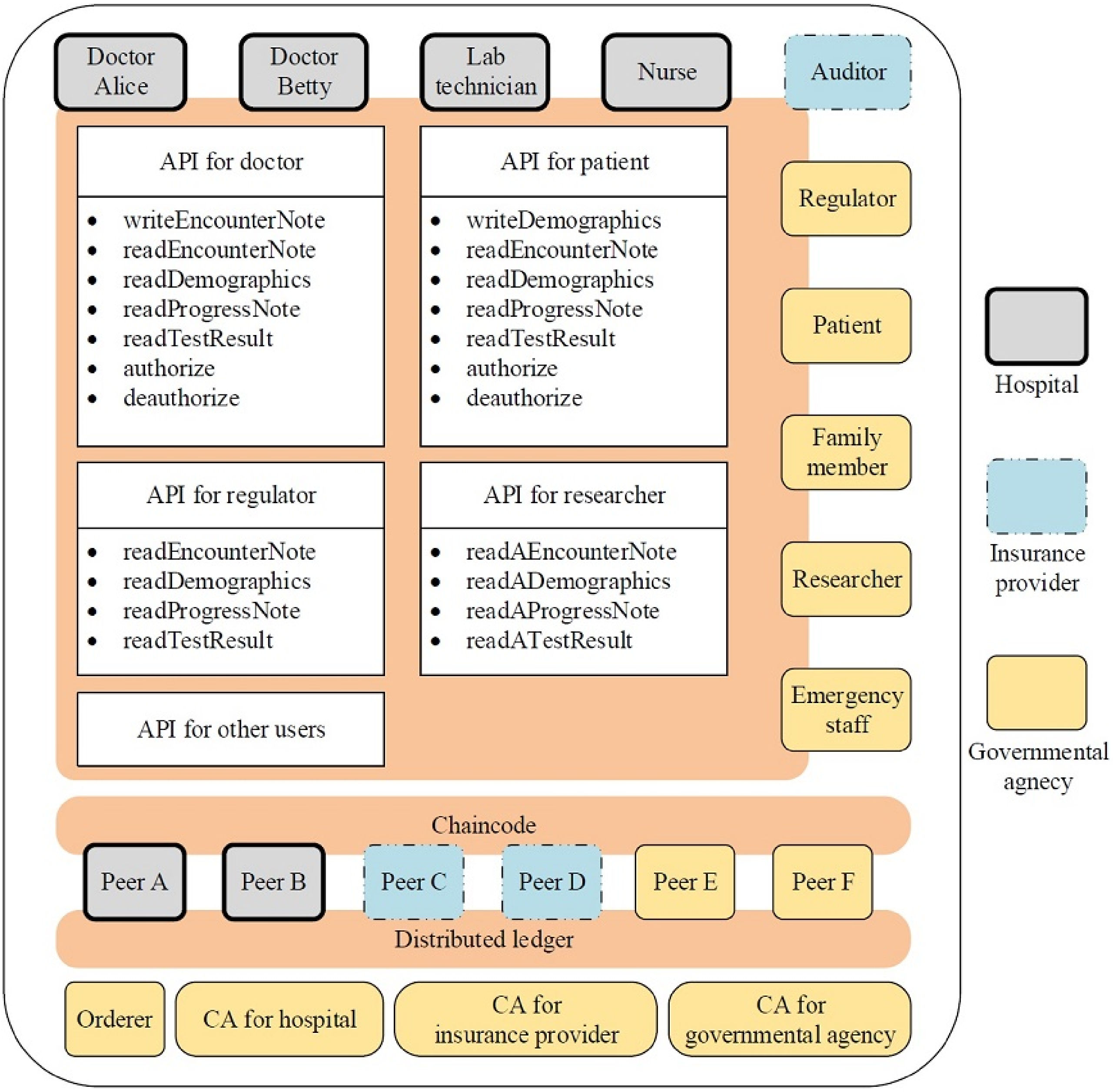
In this paper, we build a consortium blockchain for EHRs, named HealthChain, which has advantages over traditional EHR systems and other medical blockchains recently proposed. Moreover, we evaluate the performance of HealthChain through blockchain-specific metrics.

Methods

Diagram of HealthChain

The consortium consists of the following three kinds of organizations: hospitals, insurance providers, and governmental agencies. Our experimental HealthChain comprises two of each type, as shown in Figure 1. Peers A and B are servers owned by two respective hospitals, and they serve client applications of the following three kinds of users: doctors (such as Alice and Betty), lab technicians, and nurses. Peers C and D are servers contributed by two respective insurance providers, and they serve clients for auditors. Peers E and F are servers belonging to two respective governmental agencies, and they serve client applications of the following five kinds of users: regulators, patients, family members, researchers, and emergency staff. Besides, governmental agencies contribute one orderer and three certificate authorities (CAs) for three kinds of organizations. Note that these servers can reside in the cloud, in the data center of the organizations, or on a single machine.

Figure 1. The diagram of HealthChain. Three kinds of organizations contribute nodes to the blockchain network. API: application programming interface; CA: certificate authority.



Although there is no centralized node in the HealthChain network, the organizations come together under the governance model, which regulates the behavior of all actors such as peers and users. The governance model reflects the business logic of health care in real scenarios. It specifies the way that users interact with the ledger, and enables the privacy and confidentiality of EHRs. HealthChain implements the governance model via the combination of a membership service provider (MSP), chaincode, and consensus protocol. These components are explained below, and we introduce the roles of all nodes constituting the network.

CAs are trusted authorities, generating certificates and key material for actors. The MSP is implemented by these CAs, as the generated certificates can provide information about valid identities for an organization. Moreover, all actors in HealthChain can be verified by each other, and the MSP helps achieve fine-grained access and trace behavior of actors.

Peers make up the physical structure of our network. The distributed ledger and chaincode are shared among them, as illustrated in Figure 1. They are both endorsing and committing peers in HealthChain. As endorsing peers, they are designated by the consortium to execute chaincode in simulation. The chaincode implements APIs, which are divided into different groups granted to different users [32], that is, the chaincode specifies who has what access permissions to which part of the ledger and implements the fine-grained access required by law [1]. In Figure 1, we list APIs for doctors, patients, regulators, and researchers.

As committing peers, they validate blocks and commit them to their copies of the ledger. Most importantly, they check the identity of who executes the EHR request and the identity of who packages EHR transactions into a block. This is why the consensus protocol adopted in HealthChain is called PoA.

Orderers are trusted by all organizations, and they are responsible for ordering EHRs into a block on a first-come-first-serve basis. The signature of the block writer (ie, the orderer) is contained in the block. Prior to the commitment, the peers must make sure that the signature comes from an authenticated and authorized orderer, that is, the validity of a block depends on the identity of the orderer.

Identities and Anonymization

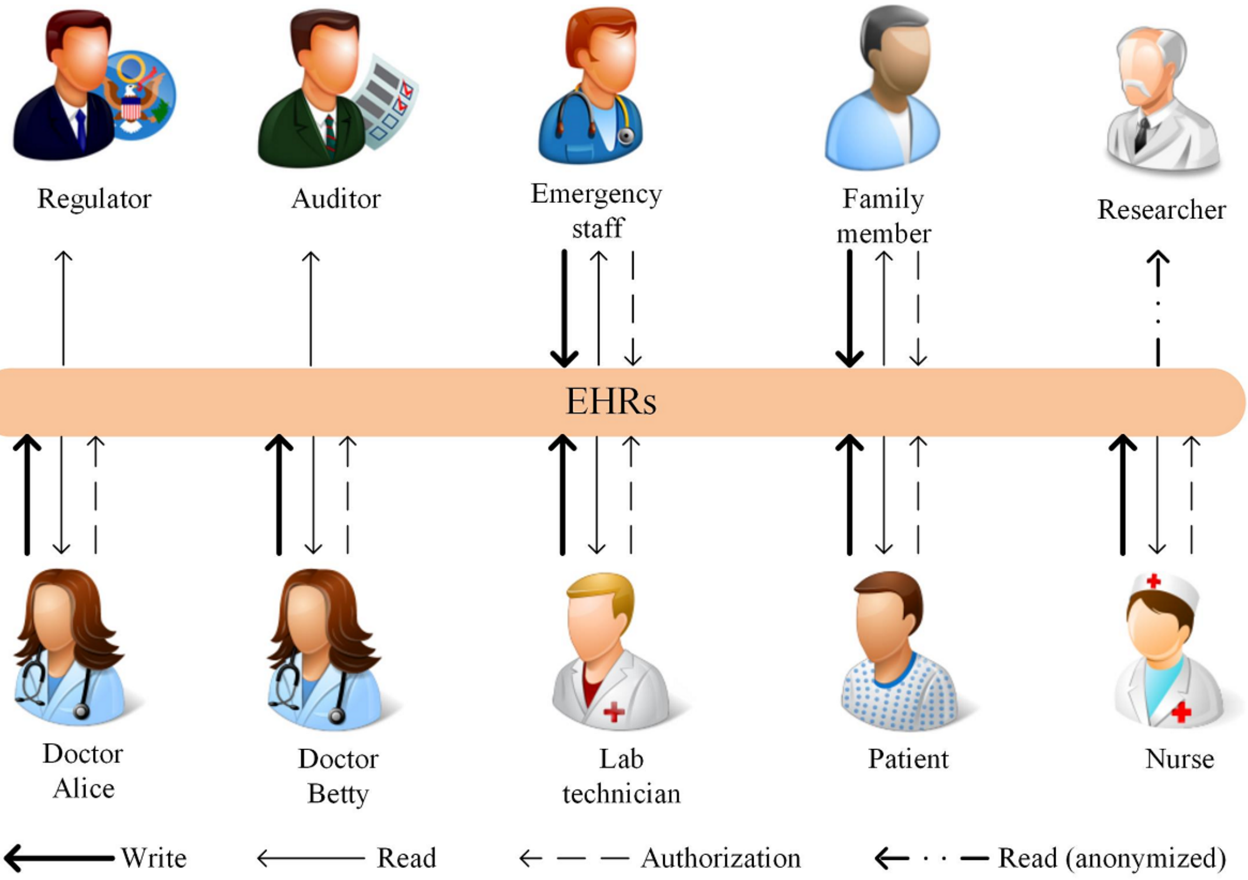
CAs dispense X.509 certificates to identify servers and clients. X.509 certificates are used in the lifecycle of transactions. For example, the EHR request contains the certificate of a client. Meanwhile, the EHR response includes the certificate of an endorsing server. When enforcing the access control of HealthChain, the chaincode extracts the certificate from the transaction request, acquires the identity of the client, and queries whether the access is authorized.

Considering privacy preservation, the identities of patients and health care providers need to be anonymized when researchers access the ledger via the APIs for them. During the anonymization, we use the hash of identities instead of the identities themselves. Thus, the identities are kept private, but the relationship between patients and health care providers is retained. Specifically, SHA256 is adopted as it is the default hash algorithm in Hyperledger Fabric. The hash of the patient ID combined with the timestamp is calculated. The purpose of introducing the timestamp is to prevent getting the same hash among queries.

Access Rights of Users

HealthChain is a distributed and append-only ledger shared among many users with different access rights. To achieve a delicate balance between privacy and availability, fine-grained access to the ledger is implemented. Figure 2 illustrates the following access rights of users in HealthChain: write permission, read permission, authorization permission, and read permission with EHRs anonymized.

Figure 2. The access rights of users in HealthChain. Different users have different access qualifications for electronic health records (EHRs).



Patients manage and control access to the PHRs, conceptually including patient-reported information and EHRs. The former includes various contents, such as demographics, allergies, and monitoring data collected from instruments [5]. The latter refers to medical records updated by doctors and staff. Patients can authorize their family members or health care providers to write and read their health information [35], reducing the risk of data replication and tracking possible trends and changes in their health.

Doctors, nurses, emergency staff, and lab technicians manage and control access to the EHRs updated by themselves. Furthermore, they can use or disclose protected health information for treatment, payment, or health care operations without patients' authorization [36]. Therefore, they have authorization permission to grant write or read permission to other covered entities, whereby the EHRs are shared across health care organizations.

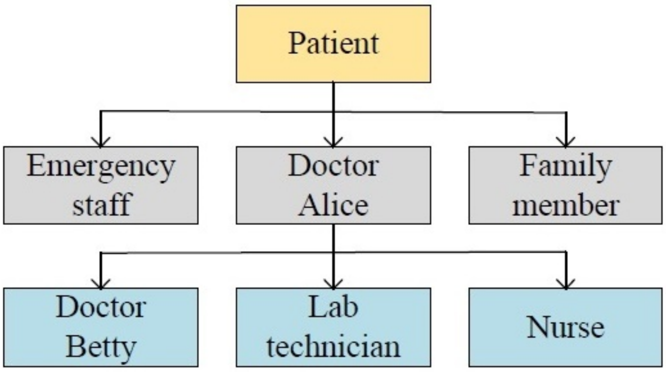
Regulators and insurance providers only have read permission to the ledger. Regulators, such as the Department of Health & Human Services, ensure that the business logic of HealthChain has been respected well, and all participants behave appropriately. When there is a dispute, they can make decisions based on the ledger, which is tamper-resistant and unforgeable [37]. Insurance providers process medical claims and evaluate their validity according to the records on HealthChain.

Researchers are engaged in public health activities such as disease surveillance. HealthChain can provide trustworthy data for this purpose because of the transparency of the data aggregation process [32]. However, the data on the ledger should be anonymized before being used for privacy preservation.

Implementation of the Authorization

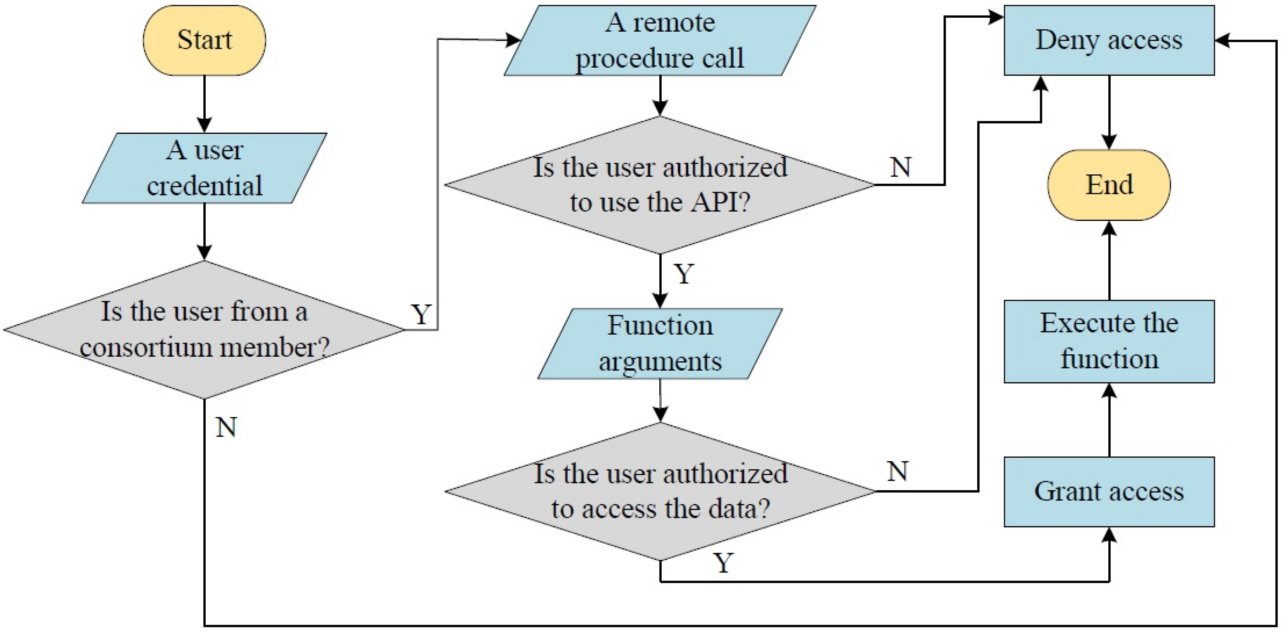
The authorization is represented by a tree data structure, as shown in Figure 3. All access rights are authorized from the patient (ie, the root of the tree). Parent nodes grant child nodes permissions such as read, write, and authorization. For example, the patient grants all three permissions to doctor Alice, and doctor Alice grants all three permissions to doctor Betty, and so on. Besides, the other three roles (ie, regulator, auditor, and researcher) have only read access to the ledger. While the tree data structure is the logical design of the authorization, the records or transactions on the ledger are its physical implementation.

Figure 3. The authorization hierarchy in HealthChain. The arrows denote the authorization operation.



The algorithm addressing authentication and authorization is presented in Figure 4, and it is implemented by the chaincode. The key parts are the three decision symbols, representing the prerequisites for the success of transactions. First, only consortium members can access the ledger; others have no right at all. Second, different types of users have different access permissions for chaincode APIs, even in the same consortium. Third, a user cannot access the data of another if the former is not authorized by the latter. For example, if a person named Carl would like to submit an encounter note about a patient named Steve, he needs to satisfy the following three conditions: he is from a hospital within the consortium, he is a doctor, and he has authorization from Steve.

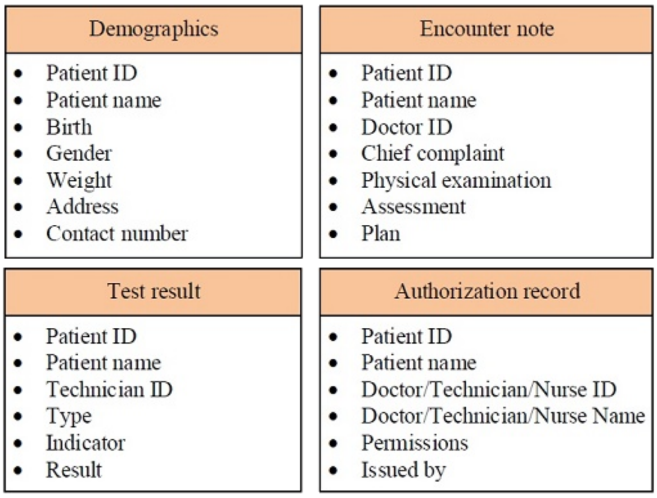
Figure 4. The fine-grained access control implemented by the chaincode. N: no; Y: yes.



Types and Lifecycle of EHRs

There are different kinds of EHRs submitted by different users. For the sake of brevity, Figure 5 lists only four of them, illustrating the data structures of the demographics, encounter note, test result, and authorization record. For example, the encounter note includes the following fields: patient ID, patient name, doctor ID, chief complaint, physical examination, assessment, and plan.

Figure 5. The types of electronic health records. Only four examples are listed here.



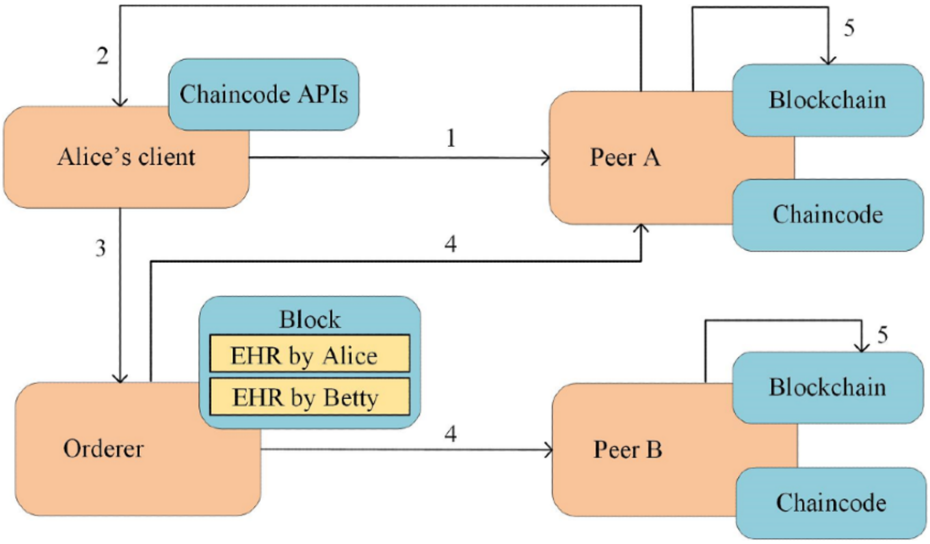
Demographics	Encounter note
Patient ID Patient name Birth Gender Weight Address Contact number	Patient ID Patient name Doctor ID Chief complaint Physical examination Assessment Plan

Test result	Authorization record
Patient ID Patient name Technician ID Type Indicator Result	Patient ID Patient name Doctor/Technician/Nurse ID Doctor/Technician/Nurse Name Permissions Issued by

Every EHR corresponds to a transaction, which needs to be executed and eventually included in the ledger. All EHRs have the same lifecycle [16], which is illustrated in Figure 6 and is explained as follows:

1. When Doctor Alice needs to record an encounter note on the ledger, her client application sends an EHR request to the endorsing peer. The request is formatted as a remote procedure call through chaincode APIs [38].
2. The endorsing peer checks the validity of the incoming request (the format, the signature, and the access permission). Thereafter, the request is processed by the chaincode, which outputs an EHR response, including returned value, read set, and write set.
3. Alice's client checks the signature of the incoming response, assembles the response and the signature into an EHR transaction, and sends the transaction to an orderer.
4. The orderer simply receives transactions from all clients (including Alice's client), orders them chronologically, and creates a block of transactions. Thereafter, the orderer delivers the block to all peers in the network.
5. Every peer checks the validity of the incoming block, including the signature and the version number. After that, the block is committed to the ledger and so is the EHR submitted by Alice.

Figure 6. The lifecycle of electronic health records (EHRs). They start with the request from a client and end with the commitment on all peers. API: application programming interface.



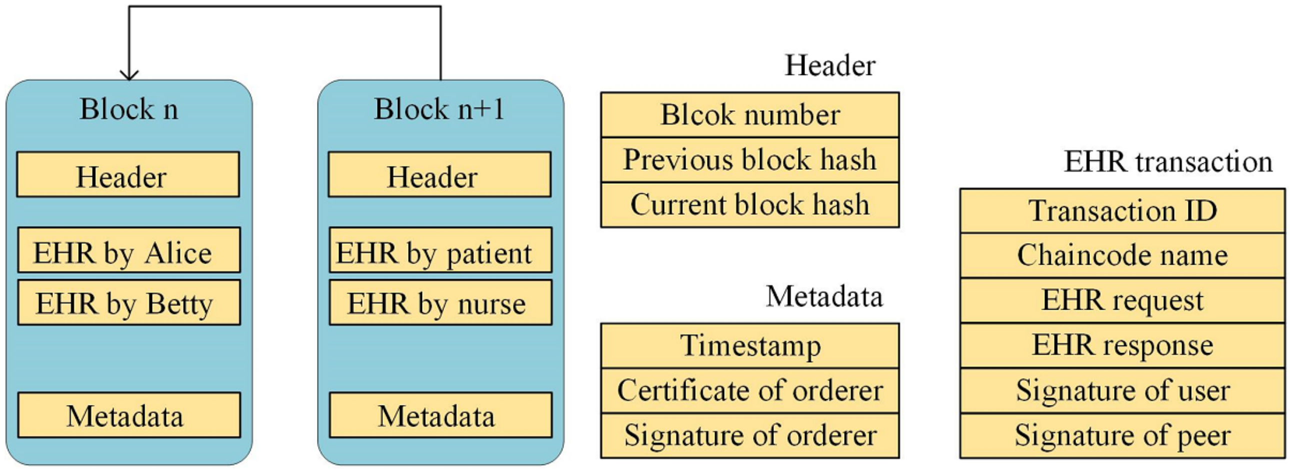
Structure of the Ledger

The ledger is the database of EHRs across time and space. As shown in Figure 7, it is structured as interlinked blocks, each of which contains chronologically ordered EHR transactions.

A block consists of the following three sections: header, EHR transactions, and metadata. The header comprises the block number, the previous block hash, and the current block hash. The EHR transactions are submitted by users such as doctors and patients. The metadata comprises the timestamp, as well as the certificate and signature of an orderer. Besides, an EHR transaction consists of the following six fields: the transaction ID, the chaincode name, the EHR request, the EHR response, and the signatures of the user and peer.

The previous block hash of block n+1 is equal to the current block hash of block n, such that the blocks are interlinked; hence, the name blockchain. The biggest difference between HealthChain and other blockchains is that blocks do not need to contain a nonce field to achieve the given pattern of cryptographic hash values. The validity of the block is only dependent on PoA, instead of proof of work. Thus, HealthChain can achieve low transaction latency.

Figure 7. The structure of the transaction, the block, and the ledger. EHR: electronic health record.



Results

Experimental Environment

The prototype of HealthChain is implemented with Hyperledger Fabric v1.4.1, an enterprise-grade permissioned distributed ledger platform [16]. It is deployed on a machine with an Intel Xeon E5-26xx v4 2.4 GHz CPU and 2 GB RAM running Ubuntu 16.04.1 LTS. All servers are built with Docker 18.06.1-ce, that is, all peers and orderers are virtualized into containers sharing the hardware and the operating system kernel [39]. The HealthChain network is created by Docker Compose, a tool for defining and running multicontainer Docker applications [40].

Besides, the two parameters of batch timeout and batch size have a high impact on the performance of HealthChain. The former denotes the maximum time to wait before creating a block, and the latter is the maximum number of EHR transactions in a block. No matter which one is satisfied first, the block is generated. This paper tests the performance of HealthChain with the parameters varying. We ran every transaction three times, and the average values of latency and throughput are provided in the paper.

Example of the Working of HealthChain

In this section, we present an example to illustrate how HealthChain works. Note that all tasks are completed through our chaincode APIs described in Figure 1. The medical data are extracted from a previous report [41].

First, the patient client of Steve Apple prepares personal information. Consequently, the following demographic record is committed to the ledger.

Second, when Steve decides to see a doctor named Carl Savem, he submits an authorization record to grant permissions for health care. Thus, the doctor is able to perform read, write, and even authorization operations on the ledger.

Third, after knowing Steve’s feelings about the health state and doing a medical examination, Carl enters the diagnosis and the instruction for treatment.

Fourth, to check whether there is a change in the lipid profile, Carl also writes an authorization record to order a test provided by a technician named John Doe.

Finally, John Doe puts the following report on the ledger.

Besides, when researchers read the above encounter note, the anonymization scheme in HealthChain takes effect, resulting in the following record. The first three items are the hashes of corresponding items in the original record and timestamps.

Security Test

We test the security of HealthChain from three aspects. First, a patient tries to access the ledger on an unavailable server, peer E. The below command queries the encounter notes about the patient with ID equal to P01, but it fails due to the unavailability of peer E, whose domain name is peer0.org3.health.com. After peer0.org3.health.com is replaced with peer0.org1.health.com, that is, the domain name of peer A, the command runs successfully since peer A is still available.

Second, an invalid user tries to access the ledger. However, this operation is immediately denied by the system since the identity cannot be authenticated. After we delete the digital certificate of Steve Apple, the above command fails, and the below information is returned.

Third, a doctor tries to tamper with an EHR. However, the history of the EHR is recorded on the ledger. We can query the history of the encounter note about P01. It is easy to figure out that “physicalExamination” has changed from “no acute distress” to “acute distress.”

In conclusion, HealthChain performs well in several aspects, namely data availability, access control, and data integrity.

Test Scenario: A Single Read

The read operation refers to retrieving or querying EHRs, and there is no change to the ledger. Users may read the EHRs submitted by themselves or others. We can give some use cases here. Doctors query the demographics of a patient or the diagnoses from former doctors. Patients read medical notes from doctors or lab results from lab technicians [24]. Insurance providers inspect EHRs to validate the necessity of them. Read latency is the time between when the EHR read request is submitted and when the EHR response is returned, corresponding to steps 1 and 2 in Figure 6.

Table 1 shows the example data of a single read operation under different parameter conditions. It can be seen that HealthChain has the read latency of about 0.1 s irrespective of the parameters. For example, the first row shows the process of a read operation when Doctor Alice retrieves an EHR from peer A. She sends an EHR request at 0 s, peer A finishes retrieving at 0.136 s, and an EHR response is returned to her at 0.142 s. Therefore, the read latency of 0.142 s is obtained with batch timeout equal to 20 s and batch size equal to 1000. The two parameters are irrelevant since there is no block generated during the read process.

Table 1. Example read latency data at different nodes.

Number	Batch timeout (s)	Batch size (n)	Client	Sever	Read request (s)	Read (s)	Response (s)	Read latency (s)
1	20	1000	Doctor	Peer A	0	0.136	0.142	0.142
2	20	1	Insurance provider	Peer C	0	0.117	0.122	0.122
3	2	10	Patient	Peer E	0	0.099	0.105	0.105

Test Scenario: Concurrent Reads

Usually, many users read from the ledger simultaneously and the read operations are executed in overlapping time periods. For example, doctors, patients, and insurance providers may read EHRs from the ledger concurrently in accordance with their needs. Read throughput is obtained by the number of read operations completed in a specific time period, indicated as reads per second (RPS).

Table 2 shows the example data of 1000 concurrent reads under different parameter conditions. It can be observed that every server of HealthChain has a read throughput of around 10 RPS in spite of varying parameters. For example, the first row shows the process of 1000 concurrent read operations when users try to retrieve EHRs from peer A at the same time. The first EHR request happens at 0 s, peer A finishes the 1000th read at 97.513 s, and the 1000th response happens at 97.521 s. Therefore, the read throughput reaches 10.259 RPS with batch timeout of 20 s and batch size of 1000. The two parameters are irrelevant as no block is written on the ledger. Furthermore, the read throughput of the whole network is the summation of the throughput of all peers as the read operations on one peer are independent of those on another.

Table 2. Example read throughput data at different nodes.

Number	Batch timeout (s)	Batch size (n)	Sever	1st read Request (s)	1000th read (s)	1000th response (s)	Read latency (s)	Read throughput (reads per s)
1	20	1000	Peer A	0	97.513	97.521	97.521	10.259
2	20	1	Peer C	0	95.847	95.855	95.855	10.440
3	2	10	Peer E	0	98.871	98.876	98.876	10.115

Test Scenario: A Single Write

The write operation refers to submitting EHRs to the ledger, and there are changes involved. Users may create and submit EHRs regarding a patient. We can list some use cases here. Doctors create encounter notes after meeting with patients or progress notes during the course of a hospitalization. Patients update demographics or report their clinical status [22]. Lab technicians report test results. Transaction latency is the time between when an EHR write request is submitted and when the EHR transaction is widely available in the network, corresponding to all five steps in Figure 6.

Table 3 shows the example data of a single write operation with different parameter conditions. It can be seen that the transaction latency depends on the two parameters. For example, the first row shows the process of a write operation when doctor Alice submits an EHR to the ledger. She sends an EHR request at 0 s, peer A endorses the EHR at 0.118 s, and the orderer generates a block containing the EHR transaction at 20.166 s. Subsequently, six peers separately commit the block to the ledger at 20.370 s, 20.352 s, 20.379 s, 20.379 s, 20.363 s, and 20.379 s. Therefore, the EHR transaction is available on all peers at the latest time, namely 20.379 s. Consequently, the transaction latency of 20.379 s is obtained with batch timeout of 20 s and batch size of 1000.

Table 3. Example transaction latency data at different nodes.

No.	Batch timeout (s)	Batch size (n)	Client	Write request (s)	Endorsement on A (s)	Write on orderer (s)	Commitment on A (s)	Commitment on B (s)	Commitment on C (s)	Commitment on D (s)	Commitment on E (s)	Commitment on F (s)	Latest (s)	Transaction latency (s)
1	20	1000	Doctor	0	0.118	20.166	20.370	20.352	20.379	20.379	20.363	20.379	20.379	20.379
2	20	1	Patient	0	0.110	0.163	0.434	0.424	0.423	0.428	0.437	0.418	0.437	0.437
3	2	10	Lab technician	0	0.099	2.141	2.292	2.306	2.290	2.305	2.306	2.291	2.306	2.306

We explain the transaction latency in Table 3. In the first and third cases, the orderer has to wait for 20 s and 2 s, respectively, before creating a block because there is only one incoming EHR transaction that needs to be packaged into the block and the batch timeout occurs first. In the second case, the batch size is 1 and there happens to be one EHR transaction, so the batch size is satisfied first, and the orderer does not have to wait for 20 s before creating the block.

Test Scenario: Concurrent Writes

Usually, many users write to the ledger simultaneously and the write operations are executed in overlapping time periods. For example, doctors, patients, and lab technicians may write EHRs to the ledger concurrently as needed. Transaction throughput is calculated by the number of EHR transactions committed by the network in a specific time period, expressed as transactions per second (TPS).

Table 4 shows the example data of 1000 concurrent writes with different parameter conditions. It can be seen that the transaction throughput is determined by the two parameters. For example, the first row shows the process of 1000 concurrent write operations when users try to submit EHRs to the ledger at the same time. The first EHR request occurs at 0 s, peer A finishes the 1000th endorsement at 132.224 s, and the orderer generates the last block at 139.930 s. Subsequently, six peers of the network separately commit the last block at 141.498 s, 141.936 s, 141.933 s, 141.400 s, 141.042 s, and 141.976 s. Therefore, the 1000 EHR transactions are available on all peers at the latest time, namely 141.976 s. Consequently, the transaction throughput reaches 7.043 TPS with batch timeout of 20 s and batch size of 1000.

Table 4. Example transaction throughput data at different nodes.

Number	Batch timeout (s)	Batch size (n)	1st write request (s)	1000th endorsement on A (s)	Last write on orderer (s)	Last commitment on A (s)	Last commitment on B (s)	Last commitment on C (s)	Last commitment on D (s)	Last commitment on E (s)	Last commitment on F (s)	Latest (s)	Transaction latency (s)	Transaction throughput (TPS)
1	20	1000	0	132.224	139.930	141.498	141.936	141.933	141.400	141.042	141.976	141.976	141.976	7.043
2	20	1	0	241.035	241.519	633.266	608.020	635.720	608.755	636.811	609.583	636.811	636.811	1.570
3	2	10	0	157.787	157.818	158.096	158.202	158.153	158.100	158.094	158.126	158.202	158.202	6.321

We can account for the transaction throughput in Table 4. In the first case, blocks are created every 20 s. Because it takes more than 20 s for 1000 transactions to come, the batch timeout happens first. In the second case, blocks are created every EHR transaction since the batch size is 1. In the third case, blocks are created every 10 EHR transactions. Because it takes less than 2 s for 10 transactions to come, the batch size happens first. It can be observed that the transaction throughput is inversely proportional to the number of blocks generated. Generating more blocks increases the likelihood of network congestion caused by the gossip protocol [16].

Discussion

Features of HealthChain

Based on the content that we have covered, the features of HealthChain are summarized as follows. First, HealthChain is permissioned. Unlike with a public permissionless network, all users like doctors and patients are certificated by the MSP and therefore are identifiable to each other, rather than anonymous and fully untrusted. Unauthorized or unknown users are not allowed to access the ledger. Second, HealthChain is immutable, that is, once EHRs have been added to the chain, they cannot be changed. This append-only property depends on the fact that the blocks are interlinked via hash references. Thus, HealthChain is the authoritative source of patients' treatment history. Third, HealthChain is transparent. Health care stakeholders come together to constitute the blockchain network, and none of them controls the whole system. Every operation initiated by users is checked against the governance model, which regulates and monitors the behavior of all actors. Furthermore, EHRs like encounter notes and lab results are shared among covered users, who know what is going on during the course of treatment. Fourth, HealthChain is HIPAA compliant with the privacy rule and security rule. Privacy policies are implemented through chaincode. The use of patient information is denied without authorization. Patients have final control over the EHRs, and they can grant write and read permissions to other covered entities. Besides, HealthChain exploits the Transport Layer Security (TLS) protocol to provide communication security over the network [16]. Fifth, HealthChain is scalable. Not all peers are involved in the transaction execution, and not all orderers are involved in the block generation. Therefore, parallel transaction execution and block generation are allowed, and HealthChain can easily support more nodes, though there are only six peers and one orderer in our experimental environment. Sixth, HealthChain has good performance. Even in our experimental environment, the read latency was about 0.1 s, and the read throughput of every peer was about 10 RPS. HealthChain achieves a transaction latency of about 0.4 s with batch size of 1, and it supports a transaction throughput of about 7 TPS with batch timeout of 20 s and batch size of 1000. In contrast, Bitcoin and Ethereum take 600 s and 10 s, respectively, to write a transaction on the ledger [42].

Parameter Setting

Performance of HealthChain is affected by many variables such as network size and limits of the hardware. Here, we discuss two parameters that we configure in the experiments, that is, batch timeout and batch size.

To achieve good performance, we should adjust batch timeout considering the permitted maximum transaction latency and set the batch size according to the rate indicating how many EHR transactions are submitted to the orderer during a specific period. The two parameters should be separately proportional to the permitted latency and the rate. As shown in the experiments, HealthChain obtains the lowest transaction latency when there is only one EHR transaction and the batch size is set to 1. It achieves the highest transaction throughput when there are 1000 transactions and the batch size is 1000.

Limitations

The prototype of HealthChain has two disadvantages. First, the only orderer causes a single point of failure. If the orderer fails, EHR transactions cannot be ordered into a block, causing failure of the entire system. To address this problem, we can deploy an ordering service consisting of a set of ordering service nodes and a Kafka cluster with its ZooKeeper ensemble [16]. This will help to not only build a crash fault-tolerant system but also increase the performance owing to load balancing.

Second, the read latency increases with the growth of the ledger. The ledger is implemented as a file on the disk considering the append-only write operation. At the same time, the read operation is also common in HealthChain. However, EHRs may scatter over the file with time, resulting in difficulty in searching for them. To solve the problem, we can create an index of patients on the ledger, making the read operation fast.

Conclusions

In this study, we built and evaluated HealthChain, which is an EHR consortium blockchain that operates under a governance model. It ensures data availability and data integrity. It provides chaincode APIs to accommodate the requirements from different clients and implements fine-grained access control. Besides, a way to anonymize EHRs is introduced. HealthChain adopts PoA as its consensus algorithm. The functionality of HealthChain was observed in the experiments. We described the performance of the system through latency and throughput.

Acknowledgments

The authors would like to acknowledge the support of Doctoral Top Talents Program of CQUPT under Grant BYJS2016004. The authors also thank the instructor Tao Wu for his comments on the paper.

Conflicts of Interest

None declared.

References

1. HITECH Act Enforcement Interim Final Rule. US Department of Health & Human Services. URL: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/enforcementrule/enfifr.pdf [accessed 2018-12-16] [WebCite Cache ID 74hLwxin3]

2. Neal D. Choosing an electronic health records system: professional liability considerations. Innov Clin Neurosci 2011 Jun;8(6):43-45 [FREE Full text] [Medline: 21779542]
3. Xhafa F, Li J, Zhao G, Li J, Chen X, Wong DS. Designing cloud-based electronic health record system with attribute-based encryption. Multimed Tools Appl 2014 Feb 11;74(10):3441-3458. [doi: 10.1007/s11042-013-1829-6]
4. McMullen PC, Howie WO, Philipsen N, Bryant VC, Setlow PD, Calhoun M, et al. Electronic Medical Records and Electronic Health Records: Overview for Nurse Practitioners. The Journal for Nurse Practitioners 2014 Oct;10(9):660-665. [doi: 10.1016/j.nurpra.2014.07.013]
5. Defining Key Health Information Technology Terms. US National Alliance for Health Information Technology. URL: http://www.nachc.org/wp-content/uploads/2016/03/Key-HIT-Terms-Definitions-Final_April_2008.pdf [accessed 2019-1-19] [WebCite Cache ID 74iBxwwKB]
6. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. Health Policy and Technology 2017 Mar;6(1):20-25. [doi: 10.1016/j.hlpt.2016.08.002]
7. Valle J, Gomes C, Godby T, Coustasse A. The Feasibility of the Nationwide Health Information Network. The Health Care Manager 2016;35(2):103-112. [doi: 10.1097/hcm.0000000000000101]
8. Lapsia V, Lamb K, Yasnoff WA. Where should electronic records for patients be stored? Int J Med Inform 2012 Dec;81(12):821-827. [doi: 10.1016/j.ijmedinf.2012.08.008] [Medline: 23021932]
9. Kierkegaard P. Medical data breaches: Notification delayed is notification denied. Computer Law & Security Review 2012 Apr;28(2):163-183. [doi: 10.1016/j.clsr.2012.01.003]
10. Summary of the HIPAA Privacy Rule. US Department of Health & Human Services. URL: https://www.hhs.gov/sites/default/files/privacysummary.pdf [accessed 2018-12-22] [WebCite Cache ID 74qceeIlV]
11. Chronology of data breaches. Privacy Rights Clearinghouse. URL: https://www.privacyrights.org/data-breaches [accessed 2018-12-22] [WebCite Cache ID 74qjpz9vL]
12. Holman T. Patient-reported data often excluded from EHR documentation. SearchHealthIT. URL: https://searchhealthit.techtarget.com/blog/Health-IT-Pulse/Patient-reported-data-often-excluded-from-EHR-documentation [accessed 2018-12-22] [WebCite Cache ID 74qqcSZY1]
13. Shae Z, Tsai J. Transform blockchain into distributed parallel computing architecture for precision medicine. 2018 Presented at: 38th International Conference on Distributed Computing Systems (ICDCS); July 2-6,2018; Vienna, Austria p. 1290-1299. [doi: 10.1109/icdcs.2018.00129]
14. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Bitcoin. URL: https://bitcoin.org/bitcoin.pdf [accessed 2018-12-23] [WebCite Cache ID 74r9hzoTN]
15. Buterin V. A next-generation smart contract and decentralized application platform. Ethereum White Paper. URL: https://cdn.relayto.com/media/files/QFTNsr6YSYCE9zyamwis_EthereumWhitePaper.pdf [accessed 2018-12-23] [WebCite Cache ID 74rB2yNtp]
16. A blockchain platform for the enterprise. Hyperledger Fabric. URL: https://hyperledger-fabric.readthedocs.io/en/latest/index.html [accessed 2018-12-23] [WebCite Cache ID 74rBVCJ3L]
17. Narayanan A, Bonneau J, Felten E. Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton, New Jersey, United States: Princeton University Press; 2016.
18. Drescher D. A Non-Technical Introduction in 25 Steps. New York, New York, United States: Apress; 2017.
19. Yli-Huumo J, Ko D, Choi S, Park S, Smolander K. Where Is Current Research on Blockchain Technology?-A Systematic Review. PLoS One 2016 Oct 3;11(10):e0163477 [FREE Full text] [doi: 10.1371/journal.pone.0163477] [Medline: 27695049]
20. Minoli D, Occhiogrosso B. Blockchain mechanisms for IoT security. Internet of Things 2018 Sep;1-2:1-13. [doi: 10.1016/j.iot.2018.05.002]
21. Azaria A, Ekblaw A, Vieira T. MedRec: Using Blockchain for Medical Data Access and Permission Management. 2016 Presented at: 2nd International Conference on Open and Big Data (OBD); August 22-24, 2016; Vienna, Austria p. 25-30. [doi: 10.1109/obd.2016.11]
22. Rifi N, Rachkidi E, Agoulmine N, Taher NC. Towards using blockchain technology for eHealth data access management. 2017 Presented at: Fourth International Conference on Advances in Biomedical Engineering (ICABME); October 19-21, 2017; Beirut, Lebanon. [doi: 10.1109/ICABME.2017.8167555]
23. Hyperledger Architecture, Volume II: Smart Contracts. Hyperledger Architecture Working Group. URL: https://www.hyperledger.org/wp-content/uploads/2018/04/Hyperledger_Arch_WG_Paper_2_SmartContracts.pdf [accessed 2018-12-23] [WebCite Cache ID 74sZtbY0P]
24. Balsari S, Fortenko A, Blaya JA, Gropper A, Jayaram M, Matthan R, et al. Reimagining Health Data Exchange: An Application Programming Interface-Enabled Roadmap for India. J Med Internet Res 2018 Jul 13;20(7):e10725 [FREE Full text] [doi: 10.2196/10725] [Medline: 30006325]
25. Buterin V. On public and private blockchains. Ethereum Foundation. 2015. URL: https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/ [accessed 2018-12-24] [WebCite Cache ID 74smalaQg]
26. Irving G, Holden J. How blockchain-timestamped protocols could improve the trustworthiness of medical science. F1000Res 2016 Feb 26;5:222. [doi: 10.12688/f1000research.8114.1]

27. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]
28. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain. IEEE Access 2017;5:14757-14767. [doi: 10.1109/ACCESS.2017.2730843]
29. Magyar G. Blockchain: Solving the privacy and research availability tradeoff for EHR data: A new disruptive technology in health data management. 2017 Presented at: 30th Neumann Colloquium (NC); November 24-25, 2017; Budapest, Hungary p. 135-140. [doi: 10.1109/nc.2017.8263269]
30. Guo R, Shi H, Zhao Q, Zheng D. Secure Attribute-Based Signature Scheme With Multiple Authorities for Blockchain in Electronic Health Records Systems. IEEE Access 2018;6:11676-11686. [doi: 10.1109/ACCESS.2018.2801266]
31. Ichikawa D, Kashiyama M, Ueno T. Tamper-Resistant Mobile Health Using Blockchain Technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]
32. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and Trustable Electronic Medical Records Sharing using Blockchain. AMIA Annu Symp Proc 2017;2017:650-659 [FREE Full text] [Medline: 29854130]
33. Sun Y, Zhang R, Wang X, Gao K, Liu L. A Decentralizing Attribute-Based Signature for Healthcare Blockchain. 2018 Presented at: 27th International Conference on Computer Communication and Networks, ICCCN; July 30-August 2, 2018; Hangzhou, China. [doi: 10.1109/ICCCN.2018.8487349]
34. Hyperledger blockchain performance metrics. Hyperledger Performance and Scale Working Group. URL: https://www.hyperledger.org/resources/publications/blockchain-performance-metrics [accessed 2018-12-25] [WebCite Cache ID 74vaMkIc5]
35. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustainable Cities and Society 2018 May;39:283-297. [doi: 10.1016/j.scs.2018.02.014]
36. Uses and disclosures to carry out treatment, payment, or health care operations. US Department of Health & Human Services. URL: https://www.govinfo.gov/content/pkg/CFR-2003-title45-vol1/xml/CFR-2003-title45-vol1-sec164-506.xml [accessed 2018-12-26] [WebCite Cache ID 74wzCpzpw]
37. Mettler M. Blockchain technology in healthcare: The revolution starts here. 2016 Presented at: 18th International Conference on e-Health Networking, Applications and Services (Health-com); September 14-16, 2016; Munich, Germany. [doi: 10.1109/healthcom.2016.7749510]
38. What is gRPC? Google Inc. URL: https://grpc.io/docs/guides/ [accessed 2018-12-27] [WebCite Cache ID 74yI2lSr8]
39. Get started with Docker. Docker Inc. URL: https://docs.docker.com/get-started/ [accessed 2019-01-01] [WebCite Cache ID 755tFUmwz]
40. Overview of Docker Compose. Docker Inc. URL: https://docs.docker.com/compose/overview/ [accessed 2019-01-01] [WebCite Cache ID 755scQ10H]
41. Sample Medical Record: Steve Apple. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/ncepcr/tools/pf-handbook/mod8-app-b-steve-apple.html [accessed 2019-06-22]
42. Gao F, Zhu L, Shen M, Sharif K, Wan Z, Ren K. A Blockchain-Based Privacy-Preserving Payment Mechanism for Vehicle-to-Grid Networks. IEEE Network 2018 Nov;32(6):184-192. [doi: 10.1109/MNET.2018.1700269]

Abbreviations

API: application programming interface
CA: certificate authority
EHR: electronic health record
EMR: electronic medical record
HIPAA: Health Insurance Portability and Accountability Act
MSP: membership service provider
PBFT: Practical Byzantine Fault Tolerance
PHR: personal health record
PoA: proof of authority
RPS: reads per second
TPS: transactions per second

Edited by K Clauson, P Zhang; submitted 31.01.19; peer-reviewed by Q Xia, F Xhafa, D Zheng, T Ueno; comments to author 27.04.19; revised version received 23.09.19; accepted 08.12.19; published 22.01.21.

Please cite as:
Xiao Y, Xu B, Jiang W, Wu Y
The HealthChain Blockchain for Electronic Health Records: Development Study
J Med Internet Res 2021;23(1):e13556
URL: http://www.jmir.org/2021/1/e13556/
doi:10.2196/13556
PMID:33480851

Corrigenda

Correction: A Newly Developed Web-Based Resource on Genetic Eye Disorders for Users With Visual Impairment (Gene.Vision): Usability Study

Jian Lee Yeong[1*], MBChB; Peter Thomas[2*], MA, BM, BCS, FRCOphth, PhD; James Buller[3], BSc; Mariya Moosajee[1,2,4], BSc, MBBS, PhD, FRCOphth

[1]Institute of Ophthalmology, University College London, London, United Kingdom
[2]Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom
[3]Aniridia Network, Sheffield, United Kingdom
[4]Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Mariya Moosajee, BSc, MBBS, PhD, FRCOphth
Institute of Ophthalmology
University College London
11-43 Bath Street
London, EC1V 9EL
United Kingdom
Phone: 1 2076086800
Email: m.moosajee@ucl.ac.uk

Related Article:

Correction of: http://www.jmir.org/2021/1/e19151/

(J Med Internet Res 2021;23(1):e27330) doi:10.2196/27330

In “A Newly Developed Web-Based Resource on Genetic Eye Disorders for Users With Visual Impairment (Gene.Vision): Usability Study” (J Med Internet Res 2021;23(1):e19151) the authors noted one error.

In the originally published article, additional text was inadvertently included at end of the Background section of the Abstract. It read as follows:

> *Despite the introduction of the Web Content Accessibility Guidelines and legislations, many websites remain poorly accessible to users with disability, especially those with visual impairment, as the internet has become a more visually complex environment. With increasing reliance on the internet and almost 2 million people in the United Kingdom being affected by vision loss, it is important that they are not overlooked when developing web-based materials. A significant proportion of those affected have irreversible vision loss due to rare genetic eye disorders, and many of them use the internet as a primary source of information for their conditions. However, access to high-quality web-based health information with an inclusive design remains a challenge for many. We have developed a new web-based resource for genetic eye disorders called Gene.Vision that aims to provide a holistic guide for patients, relatives, and health care professionals. by sight loss, it is important that they are not overlooked when developing web-based materials. A significant proportion of those affected have irreversible sight loss due to rare genetic eye disorders, and many of them use the internet as a primary source of information for their conditions. However, access to high-quality web-based health information with an inclusive design remains a challenge for many.*

This section has been revised to:

> *Despite the introduction of the Web Content Accessibility Guidelines and legislations, many websites remain poorly accessible to users with disability, especially those with visual impairment, as the internet has become a more visually complex environment. With increasing reliance on the internet and almost 2 million people in the United Kingdom being affected by vision loss, it is important that they are not overlooked when developing web-based materials. A significant proportion of those affected have irreversible vision loss due to rare genetic eye disorders, and many of them use the internet as a primary source of information for their conditions. However, access to high-quality web-based health information with an inclusive design remains a*

> *challenge for many. We have developed a new web-based resource for genetic eye disorders called Gene.Vision that aims to provide a holistic guide for patients, relatives, and health care professionals.*

The correction will appear in the online version of the paper on the JMIR Publications website on January 25, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 21.01.21; this is a non–peer-reviewed article;accepted 21.01.21; published 25.01.21.

Please cite as:
Yeong JL, Thomas P, Buller J, Moosajee M
Correction: A Newly Developed Web-Based Resource on Genetic Eye Disorders for Users With Visual Impairment (Gene.Vision): Usability Study
J Med Internet Res 2021;23(1):e27330
URL: http://www.jmir.org/2021/1/e27330/
doi:10.2196/27330
PMID:33493127

Corrigenda and Addenda

Correction: Application of an Adaptive, Digital, Game-Based Approach for Cognitive Assessment in Multiple Sclerosis: Observational Study

Wan-Yu Hsu[1], PhD; William Rowles[1], BA; Joaquin A Anguera[1,2,3], PhD; Chao Zhao[1], MSc; Annika Anderson[1], BA; Amber Alexander[1], BSc; Simone Sacco[1], MD; Roland Henry[1], PhD; Adam Gazzaley[1,2,3,4], MD, PhD; Riley Bove[1], MD

[1]Department of Neurology, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States
[2]Neuroscape, University of California, San Francisco, San Francisco, CA, United States
[3]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, United States
[4]Department of Physiology, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Riley Bove, MD
Department of Neurology
Weill Institute for Neurosciences
University of California, San Francisco
675 Nelson Rising Lane
San Francisco, CA, 94158
United States
Phone: 1 415 595 2795
Email: riley.bove@ucsf.edu

Related Article:

Correction of: https://www.jmir.org/2021/1/e24356/

(J Med Internet Res 2021;23(1):e27440) doi:10.2196/27440

In “Application of an Adaptive, Digital, Game-Based Approach for Cognitive Assessment in Multiple Sclerosis: Observational Study” (J Med Internet Res 2021;23(1):e24356) the authors noted two errors.

In the originally published manuscript, author *Joaquin Anguera* was incorrectly named. The following name has been included in the corrected version of the manuscript: *Joaquin A Anguera*.

In the originally published manuscript, a copyright credit was not included in the caption of Figure 1. The following text has been added to the end of the Figure 1 caption in the corrected version of the manuscript:

> *Copyright © 2020-2021, Akili Interactive Labs, Inc. All rights reserved.*

The corrections will appear in the online version of the paper on the JMIR Publications website on January 27, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.01.21; this is a non–peer-reviewed article;accepted 25.01.21; published 27.01.21.

Please cite as:
Hsu WY, Rowles W, Anguera JA, Zhao C, Anderson A, Alexander A, Sacco S, Henry R, Gazzaley A, Bove R
Correction: Application of an Adaptive, Digital, Game-Based Approach for Cognitive Assessment in Multiple Sclerosis: Observational Study
J Med Internet Res 2021;23(1):e27440
URL: http://www.jmir.org/2021/1/e27440/
doi:10.2196/27440
PMID:33502997

Corrigenda and Addenda

Correction: Electronic Cognitive Screen Technology for Screening Older Adults With Dementia and Mild Cognitive Impairment in a Community Setting: Development and Validation Study

Joyce Chan[1*], MPH; Adrian Wong[1,2*], PhD; Brian Yiu[1*], BBA; Hazel Mok[1*], BSc; Patti Lam[3*], MSc; Pauline Kwan[1*]; Amany Chan[3*], MPhil; Vincent C T Mok[1,2,4*], MD; Kelvin K F Tsoi[5*], PhD; Timothy C Y Kwok[1,2,3*], MD

[1]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[2]Therese Pei Fong Chow Research Centre for Prevention of Dementia, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[3]Jockey Club Centre for Osteoporosis Care and Control, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[4]Gerald Choa Neuroscience Centre, Lui Che Woo Institute of Innovative Medicine, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[5]Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[*]all authors contributed equally

Corresponding Author:
Timothy C Y Kwok, MD
Department of Medicine and Therapeutics
The Chinese University of Hong Kong
9/F., Lui Che Woo Clinical Sciences Building, Prince of Wales Hospital
30-32 Ngan Shing St, Shatin, N.T.
Hong Kong
China (Hong Kong)
Phone: 852 3505 3145
Email: tkwok@cuhk.edu.hk

Related Article:

Correction of: https://www.jmir.org/2020/12/e17332/

(J Med Internet Res 2021;23(1):e26724) doi:10.2196/26724

In “Electronic Cognitive Screen Technology for Screening Older Adults With Dementia and Mild Cognitive Impairment in a Community Setting: Development and Validation Study” (J Med Internet Res 2020;22(12):e17332) the authors noted the need to revise the “Acknowledgments” section.

In the originally published article, the Acknowledgments section read as follows:

> *The development of EC-Screen is supported by The Hong Kong Jockey Club Charities Trust. We greatly appreciate the contributions of the Jockey Club Centre for Positive Ageing in the design and development of the EC-Screen, and Mindvivid Limited for program development of EC-Screen. We also thank Ms. Anthea Ng for her help in data collection and entry.*

This section has been revised to:

> *We thank Professor JE Morley, Saint Louis University School of Medicine, USA, for having agreed to let us adapt his Rapid Cognitive Screen test into EC-Screen for older Chinese people. We are also grateful for the funding support from the Hong Kong Jockey Club Charities trust, and for the support of the Jockey Club Centre for Positive Ageing for its contribution in design and data collection. We appreciate the support from Mindvivid Limited in software development of EC-Screen. We also thank Ms. Anthea Ng, research assistant of the Division of Neurology at The Chinese University of Hong Kong, for her help in data collection and entry.*

The correction will appear in the online version of the paper on the JMIR Publications website on January 19, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 23.12.20; this is a non–peer-reviewed article;accepted 04.01.21; published 19.01.21.

Please cite as:
Chan J, Wong A, Yiu B, Mok H, Lam P, Kwan P, Chan A, Mok VCT, Tsoi KKF, Kwok TCY
Correction: Electronic Cognitive Screen Technology for Screening Older Adults With Dementia and Mild Cognitive Impairment in a Community Setting: Development and Validation Study
J Med Internet Res 2021;23(1):e26724
URL: http://www.jmir.org/2021/1/e26724/
doi:10.2196/26724
PMID:33465038

Original Paper

Rapid COVID-19 Screening Based on Self-Reported Symptoms: Psychometric Assessment and Validation of the EPICOVID19 Short Diagnostic Scale

Luca Bastiani[1], PhD; Loredana Fortunato[1], MSc; Stefania Pieroni[1], MSc; Fabrizio Bianchi[1], PhD; Fulvio Adorni[2], MPH; Federica Prinelli[2], PhD; Andrea Giacomelli[3], MD; Gabriele Pagani[3], MD; Stefania Maggi[4], PhD; Caterina Trevisan[5], MD; Marianna Noale[2], MSc; Nithiya Jesuthasan[2], MPH; Aleksandra Sojic[2], PhD; Carla Pettenati[2], MD; Massimo Andreoni[6], MD; Raffaele Antonelli Incalzi[7], MD; Massimo Galli[3], MD; Sabrina Molinaro[1], PhD

[1]Institute of Clinical Physiology, National Research Council, Pisa, Italy
[2]Institute of Biomedical Technologies, National Research Council, Milano, Italy
[3]Infectious Diseases Unit, Department of Biomedical and Clinical Sciences Luigi Sacco, Università di Milano, Milano, Italy
[4]Institute of Neuroscience, Aging Branch, National Research Council, Padova, Italy
[5]Geriatric Unit, Department of Medicine, University of Padova, Padova, Italy
[6]Infectious Diseases Clinic, Department of System Medicine, Tor Vergata University of Rome, Rome, Italy
[7]Unit of Geriatrics, Department of Medicine, Biomedical Campus of Rome, Rome, Italy

Corresponding Author:
Sabrina Molinaro, PhD
Institute of Clinical Physiology
National Research Council
Via Moruzzi 1 Pisa
Pisa, 56127
Italy
Phone: 39 0503152094
Email: sabrina.molinaro@ifc.cnr.it

Abstract

Background: Confirmed COVID-19 cases have been registered in more than 200 countries, and as of July 28, 2020, over 16 million cases have been reported to the World Health Organization. This study was conducted during the epidemic peak of COVID-19 in Italy. The early identification of individuals with suspected COVID-19 is critical in immediately quarantining such individuals. Although surveys are widely used for identifying COVID-19 cases, outcomes, and associated risks, no validated epidemiological tool exists for surveying SARS-CoV-2 infection in the general population.

Objective: We evaluated the capability of self-reported symptoms in discriminating COVID-19 to identify individuals who need to undergo instrumental measurements. We defined and validated a method for identifying a cutoff score.

Methods: Our study is phase II of the EPICOVID19 Italian national survey, which launched in April 2020 and included a convenience sample of 201,121 adults who completed the EPICOVID19 questionnaire. The Phase II questionnaire, which focused on the results of nasopharyngeal swab (NPS) and serological tests, was mailed to all subjects who previously underwent NPS tests.

Results: Of 2703 subjects who completed the Phase II questionnaire, 694 (25.7%) were NPS positive. Of the 472 subjects who underwent the immunoglobulin G (IgG) test and 421 who underwent the immunoglobulin M test, 22.9% (108/472) and 11.6% (49/421) tested positive, respectively. Compared to NPS-negative subjects, NPS-positive subjects had a higher incidence of fever (421/694, 60.7% vs 391/2009, 19.5%; *P*<.001), loss of taste and smell (365/694, 52.6% vs 239/2009, 11.9%; *P*<.001), and cough (352/694, 50.7% vs 580/2009, 28.9%; *P*<.001). With regard to subjects who underwent serological tests, IgG-positive subjects had a higher incidence of fever (65/108, 60.2% vs 43/364, 11.8%; *P*<.001) and pain in muscles/bones/joints (73/108, 67.6% vs 71/364, 19.5%; *P*<.001) than IgG-negative subjects. An analysis of self-reported COVID-19 symptom items revealed a 1-factor solution, the EPICOVID19 diagnostic scale. The following optimal scores were identified: 1.03 for respiratory problems, 1.07 for chest pain, 0.97 for loss of taste and smell 0.97, and 1.05 for tachycardia (ie, heart palpitations). These were the most important symptoms. For adults aged 18-84 years, the cutoff score was 2.56 (sensitivity: 76.56%; specificity: 68.24%) for NPS-positive subjects and 2.59 (sensitivity: 80.37%; specificity: 80.17%) for IgG-positive subjects. For subjects aged ≥60 years, the cutoff score was 1.28, and accuracy based on the presence of IgG antibodies improved (sensitivity: 88.00%; specificity: 89.58%).

Conclusions: We developed a short diagnostic scale to detect subjects with symptoms that were potentially associated with COVID-19 from a wide population. Our results support the potential of self-reported symptoms in identifying individuals who require immediate clinical evaluations. Although these results come from the Italian pandemic period, this short diagnostic scale could be optimized and tested as a screening tool for future similar pandemics.

(J Med Internet Res 2021;23(1):e23897) doi:10.2196/23897

KEYWORDS

COVID-19; screening; diagnostic scale; validation; assessment; diagnostic; symptom; survey; algorithm

Introduction

SARS-CoV-2 has led to a global pandemic; on July 28, 2020, over 16 million cases and 650,805 deaths across more than 200 countries were reported by the World Health Organization and Johns Hopkins Center for Health Security [1,2]. Italy was the first European country to be hit hard by the COVID-19 epidemic. It was also the European country with the highest number of COVID-19 deaths recorded (ie, 24,780 as of April 27, 2020) [3]. Besides the immediate human toll, the readily acknowledged and potentially long-lasting effects of the pandemic on global economies, politics, health, and privacy policies at many levels has extended beyond the development of vaccines and treatments. The rapid spread of the COVID-19 disease and its seemingly high degree of variability in its presentation among individuals has led to a level of clinical and scientific focus that has not been previously seen. This focus has encompassed both traditionally reviewed and preprint publications and resources. Collaborative groups are being formed at the local, regional, national, and international levels to address patient data collection, aggregation, and analysis in ways that may change the way research is carried out in the future [4]. To ensure that these efforts are both effective and productive, data must be evaluated in a way that is suitable for their inclusion in these activities, while still recognizing that what we understand about COVID-19 is much less than what we do not understand [5].

Due to the far-reaching scope of the pandemic, we are already confronting (1) the need to implement individual testing at a level far above current capacities to optimize individual treatment, assess disease spread, and anticipate potential strains on health care resources and personnel [6]; (2) the need for improvements in available tests, such as nasopharyngeal swab (NPS) and antibody detection tests, (ie, improvements in accuracy, specificity, and sensitivity) to enable the reliable evaluation and interpretation of data for use in clinical care and policy decisions [7]; and (3) the need to harmonize clinical observations and definitions to support the development of guidelines and prognostic and diagnostic indicators, and to develop a comprehensive understanding of COVID-19 and critical factors that can help differentiate between different patient susceptibilities, presentations of the disease, and responses to treatment [8,9].

The use of web-based surveys can greatly enhance access to broader populations in a cost-effective manner, optimize screening for individuals who may need immediate care, and provide an approach for achieving item 3 in the previous paragraph. A cross-sectional national survey, EPICOVID19, was launched on April 13, 2020 and received more than 200,000 responses [10]. The survey, which represents phase I of this study, was promoted through social media (ie, Facebook, Twitter, Instagram, and WhatsApp), press releases, internet pages, local radio and television stations, and institutional websites that called upon volunteers to contact the study website. The inclusion criteria were as follows: age of >18 years; access to a mobile phone, computer, or tablet with internet connectivity; and on-line consent to participate in this study.

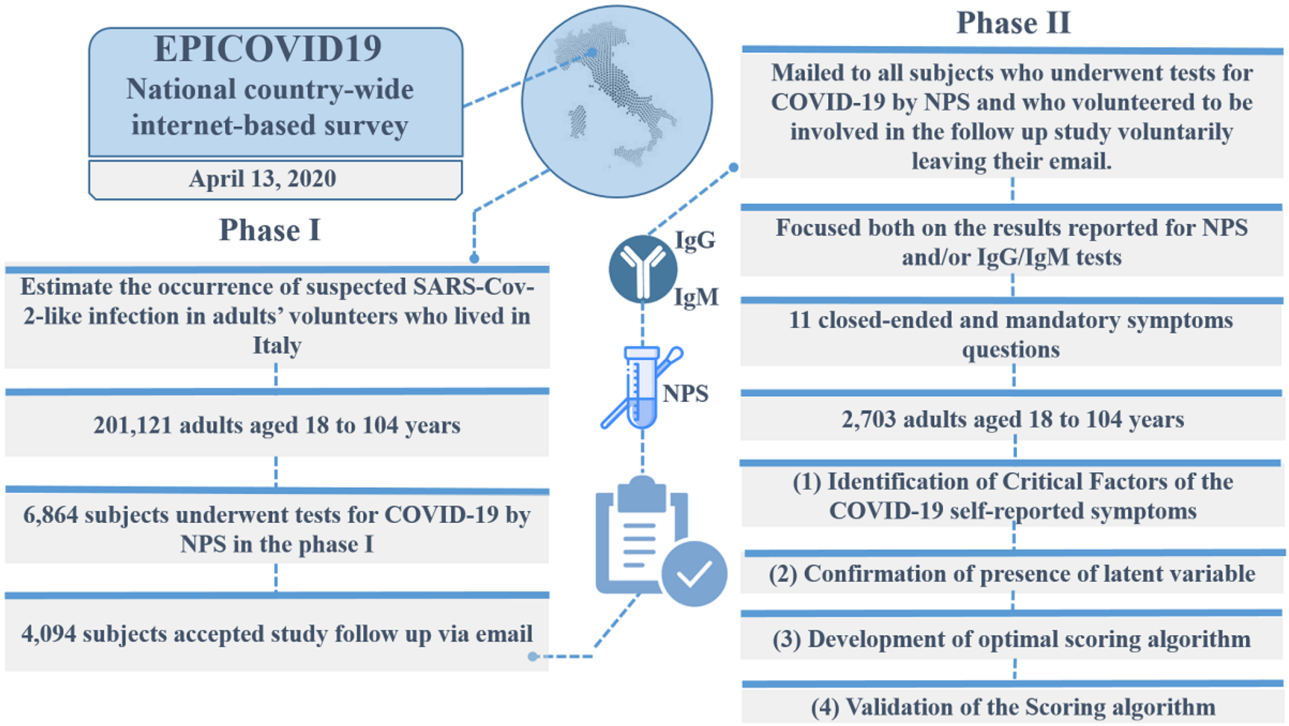
This study was conducted during the epidemic peak of COVID-19 in Italy. The aim of our study was to assess the capability of the self-reported symptoms collected through the EPICOVID19 questionnaire in discriminating COVID-19 among symptomatic subjects, in order to identify individuals with suspected COVID-19 who need to undergo instrumental measurements and clinical examinations (ie, phase II of the EPICOVID19 study). The final objectives were proposing a method for the development of a total score for the self-reported symptoms in the EPICOVID19 questionnaire, and validating the scoring method based on molecular and serological clinical diagnosis data.

Methods

Study Design and Participants

Our study is phase II of the EPICOVID19 Italian national survey [9] (pages 1-8 in Multimedia Appendix 1), which launched in April 2020 and included a convenience sample of 201,121 adults who completed the EPICOVID19 questionnaire. Figure 1 shows the overview of the EPICOVID19 2-phase study. The Phase I questionnaire investigated 6 areas through 38 questions. The 6 areas were as follows: (1) sociodemographic characteristics, (2) clinical evaluation, (3) personal characteristics and health status, (4) housing conditions, (5) lifestyle, and (6) behaviors after the lockdown.

Figure 1. Overview of the EPICOVID19 2-phase study. IgG: immunoglobulin G; IgM: immunoglobulin M; NPS: nasopharyngeal swab.



The Phase II questionnaire was mailed to all subjects who underwent NPS testing for COVID-19 and volunteered to be involved in the follow-up study in their phase I response. Phase II focused on the results of NPS and serological immunoglobulin G (IgG)/immunoglobulin M (IgM) tests and self-reported symptoms, with the aim of better identifying both symptomatic and asymptomatic SARS-CoV-2 infection cases [10].

Phase II was implemented by using an open-source statistical survey framework, LimeSurvey (version 3.17). This is a PHP (Hypertext Preprocessor)–based framework that is distributed under the GNU General Public License.

In phase II, responses to 11 questions were required. These questions covered the administration of the NPS and serological tests and the time that elapsed between observed/reported symptoms and clinical examination (ie, NPS and IgG/IgM tests) (pages 1-8 in Multimedia Appendix 1).

Of the 6864 subjects who underwent NPS testing for COVID-19 in phase I, 4094 subjects were invited by email to complete the Phase II questionnaires via the internet. Of these 4094 subjects, 38 could not participate because their email invitations were not delivered due to various issues (eg, wrong email address, full mailbox, host or domain name not found, etc), 101 refused to provide consent, and 1252 received the email, but did not proceed to complete the questionnaire.

The web-based survey included questions with close-ended answers in order to facilitate questionnaire compilation and avoid errors in digitizing answer values. At the end of the Italian lockdown period on May 2, 2020, the survey was closed and all collected data were exported for analysis with statistical tools. The base data for the statistical analysis was structured as a table that contained 1 row for each survey participant and as many columns as the collected responses. The questionnaire is available in pages 1-16 in Multimedia Appendix 1.

A total of 2703 subjects (response rate: 66%) completed the Phase II survey. After considering the 6864 subjects who underwent the NPS test in the Phase I survey, we compared the characteristics of 2703 respondents and 4161 nonrespondents. Respondents and nonrespondents to the Phase II survey appeared similar with respect to gender, age, the perception of their own health, and self-reported comorbidities. The details of the comparison between these 2 groups of subjects are included in page 9 in Multimedia Appendix 1. The resulting data of the 2703 subjects who completed the Phase II questionnaire were linked to the self-reported symptom results of the Phase I EPICOVID19 questionnaire, which included questions on the presence of 11 symptoms.

Statistical Analysis

We analyzed the self-reported symptoms that were collected in the survey to define a method for calculating a total score and validate the scoring method for serological and molecular clinical diagnoses. This was done by using 4 standard questionnaire validation steps.

The first step was the identification of critical factors. We determined the factorial structure of the COVID-19 self-reported symptom items via exploratory factor analysis (EFA), followed by confirmatory factor analysis (CFA). EFA and parallel analysis were performed to evaluate the performance of specific symptoms (ie, loadings) and define the number of factors underlying these loadings.

The second step was the confirmation of the presence of latent variables. We carried out CFA via structural equation modelling to confirm the presence of 1 latent variable (ie, factor) underlying the 11 symptoms that were chosen to identify

COVID-19. Several goodness-of-fit criteria were used, as follows: (1) standardized root mean square residual (SRSR); (2) root mean square error of approximation (RMSEA), which could not be >0.10; (3) comparative fit index (CFI); and (4) Tucker-Lewis index (TLI), which could not be <0.90.

The third step was the development of an optimal scoring algorithm. We developed an optimal scoring algorithm via homogeneity analysis by means of alternating least squares (HOMALS) and multiple correspondence analysis (MCA). Through the HOMALS procedure, we replaced specific dichotomous responses (ie, Yes/No) with categorical quantifications; the resulting score was the sum of the subject's symptom responses after they were recoded based on category quantifications.

The fourth step was the validation of the scoring algorithm. We validated the score by using an external objective criterion that was based on receiver operating characteristics analysis, in order to evaluate the performance of COVID-19 symptom scores in distinguishing symptomatic individuals in the complete sample (ie, participants aged between 18 and 84 years) and 2 specific age groups (ie, participants aged <60 years and ≥60 years). Since we aimed to discriminate COVID-19 cases, we calculated the sensitivity, specificity, and Youden index with the following 2 reference standards: (1) subjects who tested positive in the NPS tests versus subjects who tested negative in the NPS test, and (2) subjects who tested positive in the serological IgG tests versus subjects who tested negative in the IgG test. The overall predictive performance was evaluated via area under the curve (AUC) analysis.

All statistical analyses were carried out using R software (version 3.6.3), IBM SPSS 23 (IBM Corp), and Stata Statistical Software (Release 15; StataCorp LLC). The details of the performed statistical analyses are reported in pages 10-12 in Multimedia Appendix 1.

Ethical Approval

The Phase II EPICOVID19 study was approved by the Ethical Committee of the Istituto Nazionale per le Malattie Infettive, Institute for Research, Hospitalization and Healthcare Lazzaro Spallanzani as an amendment of the EPICOVID19 epidemiological study (approval number 93 in the trial register). Data transfer was safeguarded by means of password protection and encryption/decryption policies. All data were handled and stored in accordance with the European General Data Protection Regulation 2016/679 [11]. Informed consent details were accessible on the home page of the platform, and participants were asked to review these details before starting the Phase II questionnaire. The home page explained the purpose of the study, which data were to be collected, and how data were stored.

Subjects' email addresses were the personal data provided on a voluntary basis in phase I. In our study, email addresses were only used to (1) send email invitations for participating in the Phase II survey and (2) link the information related to NPS and IgG/IgM test results to the information on symptoms collected during the phase I survey. In the participation mail, subjects were able to participate by clicking on the provided link to the survey, not participate by ignoring the invitation, communicate with the authors by using the provided study-specific email address, and request the deletion of their email address from the database.

Results

Study Design and Participants

The characteristics and NPS, IgG, and IgM results of the 2703 subjects, which were supplied by those who completed the Phase II survey, are shown in Table 1. The sample predominantly consisted of women (1841/2703, 68.1%), and the average age was 49 years (SD 15.0 years) and 52 years (SD 14.1 years) for women and men, respectively. Of the 2703 respondents, 151 (5.6%) had a low educational status, 837 (31%) had a medium educational status, and 1715 (63.4%) had a high educational status. The most reported chronic condition by participants was hypertension (361/2703, 13.4%), followed by immune system diseases (266/2703, 9.8%), and depression and anxiety diseases (194/2703, 7.2%). The least frequently reported chronic symptoms were liver (21/2703, 0.8%) and kidney (22/2703, 0.8%) diseases. All the details are reported in page 13 in Multimedia Appendix 1.

Statistical Analysis

Of the 2703 subjects, 694 (25.7%) tested positive in the NPS test. Of these 694, 84 (12.1%) were asymptomatic. With regard to the subgroup of subjects who underwent serological testing, 472 underwent the IgG test and 421 underwent the IgM test, and 22.9% (108/472) and 11.6% (49/421) tested positive, respectively. Of the 108 IgG-positive subjects, 1 (0.9%) was asymptomatic. Of the 49 IgM-positive subjects, 5 (10.2%) were asymptomatic. With regard to NPS-positive subjects, the average number of days between initial symptoms and the day of swab execution was 9.3 days (SD 9.4 days; median 7 days, IQR 3-7 days). With regard to IgG-positive subjects, the average number of days between initial symptoms and the day of serological test execution was 36.1 days (SD 15.1 days; median 36.5 days, IQR 28-47 days). With regard to IgM-positive subjects, the average number of days from initial symptoms to the day of serological test execution was 26.1 days (SD 17.9 days; median 28 days, IQR 4-40 days). The incidence rate of the 11 symptoms reported by the 3 groups (ie, the NPS, IgG, IgM test groups) was similar between men and women. In the NPS-positive group, women only had a higher incidence of sore throat and cold and tachycardia (ie, heart palpitations) than men. In the IgG-positive group, men only had a higher incidence of headaches than women. In the IgM-positive group, women had a lower incidence of symptoms related to conjunctivitis than men.

The frequency of symptoms among NPS-positive subjects (Table 1) ranged from low rates of observation (eg, tachycardia [ie, heart palpitations]: 120/694, 17.3%; conjunctivitis: 111/694, 16%) to high rates of observation (eg, fever: 421/694, 60.7%; olfactory and taste disorders: 365/694, 52.6%). For all symptoms apart from headache, the incidence rates were significantly higher in NPS-positive subjects than in NPS-negative subjects (*P*<.001). With regard to the subgroup of individuals who underwent serological tests, the symptoms with a high incidence among subjects who tested positive were fever (IgG-positive group: 65/108, 60.2%; IgM-positive group: 28/49, 57.1%) and pain in muscles, bones, and joints (IgG-positive group: 73/108, 67.6%; IgM-positive group: 27/49, 55.1%). In the IgG serological test group, no significant difference was observed in the incidence of sore throat and cold symptoms (*P*=.23) between IgG-positive and IgG-negative subjects. The incidence of respiratory difficulty (*P*=.35), chest pain (*P*=.35), and gastrointestinal symptoms (*P*=.08) did not significantly differ between IgM-positive and IgM-negative subjects.

Table 1. Self-reported characteristics that were obtained from the Phase II survey and analyzed by using SARS-CoV-2 infection test results (N=2703).[a]

Variable	SARS-CoV-2 tests								
	Nasopharyngeal swab test, n=2703			Immunoglobulin G antibody test, n=472			Immunoglobulin M antibody test, n=421		
	Tested positive	Tested negative	*P* value	Tested positive	Tested negative	*P* value	Tested positive	Tested negative	*P* value
Number, n (%)	694 (25.7)	2009 (74.3)	N/A[b]	108 (22.9)	364 (77.1)	N/A	49 (11.6)	372 (88.4)	N/A
Women, n (%)	440 (63.4)	1401 (69.7)	.001	61 (56.5)	258 (70.9)	.005	25 (51)	260 (69.9)	.008
Age (years), mean (SD)	55.5 (18.06)	47.55 (12.81)	<.001	48.8 (11.74)	45.5 (11.49)	.009	50.6 (10.56)	45.8 (11.69)	.008
Answered questions on symptoms, n (%)									
Fever with a temperature of >37.5°C for at least 3 consecutive days	421 (60.7)	391 (19.5)	<.001	65 (60.2)	43 (11.8)	<.001	28 (57.1)	68 (18.3)	<.001
Cough	352 (50.7)	580 (28.9)	<.001	63 (58.3)	76 (20.9)	<.001	26 (53.1)	95 (25.5)	<.001
Sore throat and cold	232 (33.4)	756 (37.6)	.048	46 (42.6)	132 (36.3)	.233	16 (32.7)	135 (36.3)	.62
Headache	313 (45.1)	703 (35)	<.001	61 (56.5)	96 (26.4)	<.001	23 (46.9)	117 (31.5)	.03
Pain in muscles, bones, and joints	360 (51.9)	572 (28.5)	<.001	73 (67.6)	71 (19.5)	<.001	27 (55.1)	98 (26.3)	<.001
Loss of taste and smell	365 (52.6)	239 (11.9)	<.001	66 (61.1)	29 (8)	<.001	21 (42.9)	55 (14.8)	<.001
Respiratory difficulty (ie, sense of breathlessness at rest)	179 (25.8)	249 (12.4)	<.001	21 (19.4)	28 (7.7)	<.001	7 (14.3)	37 (9.9)	.35
Chest pain (ie, sternum pain)	136 (19.6)	251 (12.5)	<.001	26 (24.1)	25 (6.9)	<.001	7 (14.3)	37 (9.9)	.35
Tachycardia (ie, heart palpitations)	120 (17.3)	237 (11.8)	<.001	24 (22.2)	27 (7.4)	<.001	10 (20.4)	31 (8.3)	.007
Gastrointestinal complaints (ie, diarrhea, nausea, and vomiting)	289 (41.6)	452 (22.5)	<.001	54 (50)	65 (17.9)	<.001	17 (34.7)	87 (23.4)	.08
Conjunctivitis (ie, red eyes)	111 (16)	221 (11)	<.001	24 (22.2)	35 (9.6)	.001	11 (22.4)	40 (10.8)	.02

[a] Mean (SD) was used for continuous variables, which were analyzed with an independent 2-tailed *t* test, and n (%) was used for categorical variables, which were analyzed with a Chi-square test.

[b] N/A: not applicable.

The EFA, which involved the principal-component factors and Horn parallel analysis methods, pointed out 1 factor. Eigenvalues, descriptive indices, and goodness-of-fit indices for the cumulative percentage of explained data variability obtained through EFA are displayed in Table 2.

Principal-component factors analysis only highlighted 1 factor with an 89.9% proportion of explained variability, while the Horn parallel analysis identified 2 factors with eigenvalues of >1.0 and a 49.8% and 10.3% proportion of explained variability, respectively.

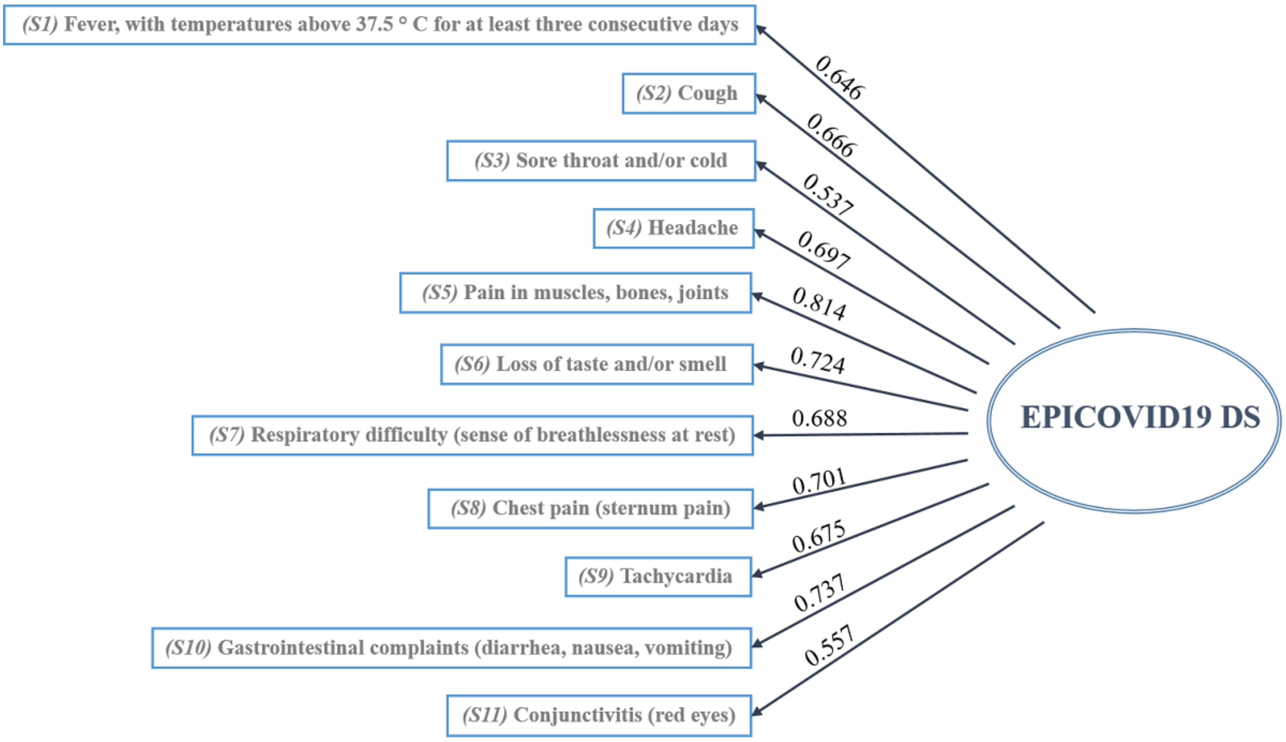
Table 2. Descriptive and goodness-of-fit dimensionality indices from the exploratory factor analysis of the 11 EPICOVID19 symptoms reported by 2703 subjects, based on the principal-component factors and Horn parallel analysis methods with an eigenvalue of >1.

Factor	Exploratory factor analysis					
	Principal-component factors analysis			Horn parallel analysis		
	Eigenvalue	Proportion of explained variability	Cumulative explained variability	Eigenvalue	Proportion of explained variability	Cumulative explained variability
1	5.00	89.9%	89.9%	5.48	49.8	49.8%
2	N/A[a]	N/A	N/A	1.14	10.3	60.1%

[a] N/A: not applicable.

Based on a priori determined cutoff value, a factor loading of >0.35 was maintained. The factor loading rule of the 1-factor solution extracted from the principal-component factors analysis is available in page 13 in Multimedia Appendix 1. The dimensionality indices of the 1-factor solution, which had a high cumulative and proportion of explained variability (89.9%), confirmed the presence of 1 latent variable underlying COVID-19 symptom items. Therefore, we defined the 1-factor solution as the EPICOVID19 diagnostic scale (EPICOVID19 DS). Based on our CFA results, we confirmed that the latent construct was unidimensional and determined how the variables contributed to the EPICOVID19 DS. Figure 2 shows the values of the standardized factor loadings for the 1-factor model. The magnitude of each factor loading value was >0.4, which indicated the importance of the corresponding item to the EPICOVID19 DS. For example, pain in muscles, bones, and joints was the most important variable, with a factor loading value of 0.814. The other variables with an optimal specific validity index were respiratory difficulty (sense of breathlessness at rest: 0.688; loss of taste and smell: 0.724) and gastrointestinal complaints, with item-factor correlations of 0.737. The lowest values were observed for the sore throat and cold and conjunctivitis items, which had a specific validity index of 0.537 and 0.557, respectively. The goodness of fit (ie, SMSR and RMSEA) of the EPICOVID19 DS was acceptable, because 2 indices were <0.10 (SMSR 0.072; RMSEA 0.052; CFI 0.977; TLI 0.971). We computed CFA indices to measure the internal validity of the model (page 14 in Multimedia Appendix 1).

Figure 2. Standardized factor loading values of the 1-factor model, EPICVOID19 DS. The goodness-of-fit indices are as follows: a standardized root mean square residual of 0.072, root mean square error of approximation of 0.052, comparative fit index of 0.977, and Tucker-Lewis index of 0.971. EPICOVID19 DS: EPICOVID19 diagnostic scale.



Given the successful unidimensionality testing of the EPICOVID19 DS, optimal scaling was performed. The proposed optimal score was extracted from the HOMALS procedure (ie, single-factor measurement), and for each subject, the computed optimal score was obtained by summing the category quantifications of the screening questionnaire item responses. Cronbach (α=0.88) and Greenacre (statistic=78%) indices confirmed the unidimensionality found in the EFA and CFA. The HOMALS optimal category quantifications of the EPICOVID19 symptom variables are summarized in Table 3, which has columns for the binary options (ie, Yes/No) and rows for the different symptoms. The HOMALS category quantifications were scaled so that the score obtained from the sum of responses would range from 0 (ie, if a subject answered “No” to all the symptoms) to 10 (ie, if a subject answered “Yes” to all the symptoms). These values are shown in the last column of Table 3. An example of a resulting score calculation is as follows: if the subject response pattern with respect to symptoms is “Yes, No, Yes, No, No, Yes, Yes, No, No, No, Yes,” the corresponding recoded response pattern is 0.80, 0, 0.64, 0, 0, 0.97, 1.03, 0, 0, 0, 0.88, and the subject’s optimal score would be calculated as 0.8 + 0 + 0.64 + 0 + 0 + 0.97 + 1.03 + 0 + 0 + 0 + 0.88 = 4.2.

Table 3. Multiple correspondence analysis optimal weights for the recoding of the EPICOVID19 diagnostic scale.

Symptoms	HOMALS[a] category quantifications		Recoded HOMALS category quantifications	
	No	Yes	No	Yes
Fever with a temperature of >37.5°C for at least 3 consecutive days	−0.362	0.8421	0	0.80
Cough	−0.426	0.810	0	0.81
Sore throat and/or cold	−0.358	0.622	0	0.64
Headache	−0.470	0.780	0	0.83
Pain in muscles, bones, and joints	−0.505	0.959	0	0.97
Loss of taste and/or smell	−0.326	1.133	0	0.97
Respiratory difficulty (ie, sense of breathlessness at rest)	−0.246	1.305	0	1.03
Chest pain (ie, sternum pain)	−0.232	1.388	0	1.07
Tachycardia (ie, heart palpitations)	−0.209	1.374	0	1.05
Gastrointestinal complaints (ie, diarrhea, nausea, and vomiting)	−0.393	1.042	0	0.95
Conjunctivitis (ie, red eyes)	−0.164	1.170	0	0.88

[a] HOMALS: homogeneity analysis by means of alternating least squares.

There was no significant difference in the mean EPICOVID19 DS score between men (mean 2.34, SD 2.2) and women (mean 2.49, SD 2.4) (*P*=.14). A low negative correlation between the scores and ages of the participants was found (ρ=−0.126; *P*<.001). Of the 2703 subjects, 1738 (64.3%) reported no preexisting diseases, 684 (25.3%) only had 1 chronic condition, while the remaining 281 (10.4%) declared ≥2 conditions. Significant differences in the mean EPICOVID19 DS score were observed between participants who did not report any disease (mean 2.26, SD 2.3) and those with at least 1 preexisting condition (mean 2.75, SD 2.4) (*P*<.001). Based on our analysis of the mean EPICOVID19 DS score among healthy subjects and subjects with 1 chronic condition, we observed significant differences between healthy subjects and subjects with lung diseases (healthy subjects: mean 2.40, SD 2.3; subjects with lung diseases: mean 3.10, SD 2.5; *P*<.001), healthy subjects and subjects with immune system diseases (healthy subjects: mean 2.39, SD 2.3; subjects with immune system diseases: mean 2.91, SD 2.4; *P*<.001), and healthy subjects and subjects with depression and anxiety diseases (healthy subjects: mean 2.42, SD 2.4; subjects with depression and anxiety: mean 2.79, SD 2.6; *P*=.036). For the other chronic conditions (ie, heart disease: *P*=.22; hypertension: *P*=.59; kidney disease: *P*=.45; tumor: *P*=.13; metabolic disease: *P*=.52; liver disease: *P*=.64), no significant differences in mean EPICOVID19 DS score were found.

The screening properties of the EPICOVID19 DS were compared to those of COVID-19–positive molecular and serological tests. These are shown in Table 4. The best Youden index value was observed for EPICOVID19 DS, with respect to subjects diagnosed with COVID-19 via NPS testing. A good trade-off between sensitivity and specificity was observed (sensitivity: 76.56%; specificity: 68.24%; AUC 77.5, 95% CI 75.6-79.4). The cutoff score obtained was 2.56. The sensitivity and specificity of the EPICOVID19 DS improved when compared to those of COVID-19–positive IgG antibody test (sensitivity: 80.37%; specificity: 80.17%; AUC 86.0, 95% CI 82.3-89.5). The cutoff value obtained (2.59) was similar to that of the NPS-positive test. The positive and negative predictive values for the IgG-positive serological test (positive predictive value [PPV]: 54.43%; negative predictive value [NPV]: 93.27%) were higher than those of the NPS test (PPV: 42.26%; NPV: 90.55%). We observed a poor performance with regard to IgM test results, so these are not presented in Table 4.

Table 4. Sensitivity and specificity of the EPICOVID19 diagnostic scale compared to those of positive COVID-19 molecular and serological diagnoses (ie, for subjects aged 18-84 years).

Statistic	SARS-CoV-2 tests	
	Nasopharyngeal swab test (n=2703), value (95% CI)[a,b]	Immunoglobulin G antibody test (n=472), value (95% CI)[c,d]
Sensitivity, %	76.56 (72.99-79.87)	80.37 (71.58-87.42)
Specificity, %	68.24 (66.16-70.28)	80.17 (75.69-84.14)
Positive likelihood ratio	2.41 (2.23-2.61)	4.05 (3.23-5.08)
Negative likelihood ratio	0.34 (0.30-0.40)	0.24 (0.17-0.36)
COVID-19–positive tests, %	23.29 (21.68-24.96)	22.77 (19.05-26.83)
Positive predictive value, %	42.26 (40.38-44.17)	54.43 (48.77-59.98)
Negative predictive value, %	90.55 (89.23-91.74)	93.27 (90.40-95.33)
Accuracy, %	70.18 (68.39-71.93)	80.21 (76.32-83.72)

[a] There were 694 NPS-positive subjects.

[b] The cutoff value for the nasopharyngeal swab test was 2.59.

[c] There were 108 immunoglobulin G-positive patients.

[d] The cutoff value for the immunoglobulin G antibody test was 2.56.

When the EPICOVID19 DS scoring algorithm was applied to specific age groups, the sensitivity and specificity of the IgG-positive antibody test (sensitivity: 88.00%; specificity: 89.58%; AUC 93.10, 95% CI 86.0-99.5) improved greatly among subjects aged ≥60 years, and the obtained cutoff value (1.28) was lower than the cutoff value for the subjects aged <60 years (2.71; sensitivity: 88.00%; specificity: 89.58%; AUC 93.10, 95% CI 86.0-99.5). The PPV and NPV of the IgG test were higher for subjects aged ≥60 years (PPV: 81.48%; NPV: 93.48%) than those for subjects aged <60 years (PPV: 51.52%; NPV: 94.38%). Furthermore, we observed the same performance in the NPS test between the specific age groups (ie, aged ≥60 years and aged <60 years), with respect to the overall sample (ie, aged 18-84 years). The details of the screening properties of the EPICOVID19 DS compared to those of COVID-19–positive molecular and serological tests for specific age groups are reported in page 16 in Multimedia Appendix 1.

Discussion

Our focus was on developing a tool composed of simple questions related to COVID-19 symptomatology for the identification of subjects who are more likely to be infected with SARS-CoV-2 in the general population. We validated the EPICOVID19 DS with a sample of voluntary subjects based on serological and molecular clinical diagnoses. The optimal score, which was computed for 2703 adults aged 18-84 years, discriminated symptomatic individuals. Before calculating the score, we performed both exploratory and confirmatory factor analyses to determine the number of factors/dimensions underlying the questionnaire. The results of these analyses supported the 1-factor model and the unidimensionality of the EPICOVID19 questionnaire. The magnitude of all factor loading values was satisfactory, and the highest factor loading values were observed for respiratory difficulty, chest pain, tachycardia (ie, heart palpitations), and loss of taste and smell. Furthermore, gastrointestinal complaint items appeared to be the most essential features of the EPICOVID19 DS. The high value for chest pain can also be explained by the fact that several patients reported it, possibly because of tracheal pain caused by pneumonia [12,13]. Several clinical studies on hospitalized patients have shown that, at the onset of COVID-19, patients frequently show typical symptoms of viral pneumonia [3]. Symptoms that are less common, but still reported by a substantial number of patients, are nasal congestion, sore throat, gastrointestinal complaints, and olfactory and taste disorders [14-16]. Subjects have often reported gastrointestinal complaints as concurrent symptoms instead of isolated symptoms of SARS-CoV-2 infection [17]. The lowest factor loading values were observed for sore throat and cold and conjunctivitis. These lower values may be related to the fact that conjunctivitis and cold are not the most frequent symptoms of COVID-19 [18]. In line with other recent studies [19,20], the features we encountered in this study showed various aspects of the definition for COVID-19 diagnosis. Cough, loss of taste and smell, and respiratory difficulty are among the most reported symptoms in previous studies, and they corresponded to the items that were the most important to our score [12,16,21,22].

The clinical presentation of COVID-19 varies, and discrepancies may exist between symptoms and the disease. A recent meta-analysis of the symptoms of 50,000 patients with COVID-19 found that fever and cough were the most common symptoms (incidence: 89.1% and 72.2%, respectively) [23], and a separate study on hospitalized subjects has suggested that respiratory distress has been reported in the most critical cases of COVID-19 [24]. With the aim of supporting medical decision making, predicted models have been developed for detecting people in the general population who are at risk of being admitted to hospital and diagnosing COVID-19 in patients with related symptoms. However, the results presented in a recent systematic review on such models describe poor research performance and a high risk of bias [25].

Based on our HOMALS, we proposed a scoring methodology for developing an improved scale. Therefore, we provided a numerical weight value (ie, optimal quantification) that represents the importance of the binary response categories (ie, Yes/No) for each question in the EPICOVID19 DS. As a result, the various binary items of the 11 questions in the EPICOVID19 DS contributed to the overall score, albeit with different weights. This produced an improved scale (ie, 0-10) that reflects the importance of each symptom. Thus, respiratory problems and chest pain were the most important symptoms, with a score of 1.03 and 1.07, respectively. The other symptoms that had an important contribution to the total score were gastrointestinal complaints (0.95), loss of taste and smell (0.97), and tachycardia (ie, heart palpitations) (1.05). Subsequently, we computed the sensitivity and specificity of EPICOVID19 DS compared to those of COVID-19–positive serological and molecular tests. For NPS-positive subjects, the cutoff score was 2.56, with a sensitivity of 76.56% and specificity of 68.24%. For IgG-positive subjects, the cutoff score was 2.59, and sensitivity, specificity, PPV, and NPV with respect to NPS-positive tests substantially improved (sensitivity: 80.37%; specificity: 80.17%; PPV: 54.43%; NPV: 93.27%). When the EPICOVID19 DS scoring algorithm was tested on subjects aged ≥60 years, the accuracy of IgG-positive antibody tests improved (sensitivity 88.00%; specificity 89.58%; AUC 93.10, 95% CI 86.0-99.5; PPV: 81.48%; NPV IgG 93.48%), and the threshold of detection (1.28) was lower than that of subjects aged <60 years.

Our data are consistent with the findings reported in previous studies. In mid-May 2020, the European all-cause mortality monitoring system showed that all-cause mortality was above the expected rate in several European countries (ie, Belgium, France, Malta, and Spain), including Italy [26], mainly for people aged ≥60 years. People aged ≥60 years are more vulnerable to SARS-CoV-2 infection, and those with preexisting medical conditions are particularly at risk. Several best practices for older people and their families have been recommended by the World Health Organization, Centers for Disease Control and Prevention, geriatricians, and infectious diseases specialists [27]. The sensitivity and specificity of serological and molecular diagnostic tests for COVID-19 have not been fully elucidated, but several studies have suggested that sensitivity could be as low as 80% [28,29]. This raises concerns of high false-negative rates, which could result in an increase in infection spread among the community. There is no absolute answer for the sensitivity and specificity of COVID-19 diagnostic tests, because to determine their accuracy, they must be compared with a gold-standard test, which does not currently exist. By considering estimates for sensitivity and specificity, PPVs and NPVs can be calculated based on disease prevalence and the rate of illness in the population. However, there is considerable uncertainty with regard to the prevalence of COVID-19 [30]. Statistically, it has been assumed that PPVs vary widely and range between 30-50% in areas with a low COVID-19 prevalence, as stated in a recent US study on COVID-19 [31].

Early recognition screening and rapid diagnosis are essential for preventing transmission and providing supportive care in a timely manner. Nevertheless, screening is different from further, more detailed diagnostic test assessments. This is of particular relevance, as resources for full testing remain limited, and optimizing the use of such resources is critical. The EPICOVID19 DS can be used as a preliminary assessment that attempts to detect subjects with symptoms that are potentially associated with COVID-19 among a wide population. The EPICOVID19 DS does not enable clinical interviews for determining complete symptomatic profiles and needs, but it does identify those who may warrant further assessment. Therefore, it would be advantageous to use the EPICOVID19 DS for screening in primary care settings, so that general practitioners can avoid people with suspected COVID-19 in primary care offices whenever possible [32]. The EPICOVID19 DS can also be used as an initial screening tool before patients are managed remotely via telephone or video consultations [33]. Additionally, the EPICOVID19 DS can be applied to the general population. Once a score is assigned to each symptom, the EPICOVID19 DS can allow for different cutoff values to be set, based on the subjects involved and the gold standards used (ie, NPS tests, serological tests, clinical evaluation by clinicians, etc).

It should be noted that since it is plausible to expect a lower prevalence rate in the general population than the 22.77% in this study, the probability of NPVs would increase beyond the current 93.27%. Consequently, the probability of progressing to COVID-19 for subjects who test negative (ie, 1 − NPV) would be less than the current 6.7%. Furthermore, although the identified symptoms in this study are not specific to COVID-19, they have been reported as valid references for a population setting, because they are frequently reported by patients with COVID-19. In a nonpandemic scenario, it is likely that these symptoms could be assessed with different weights because of their aspecificity, which would configure the EPICOVID19 DS as a valid diagnostic support tool for pandemic situations. Moreover, health authorities are still unable to use classic tests to monitor the spread of SARS-CoV-2 infection, and allowing the circulation of unsuspecting individuals with COVID-19 could represent a risk for the spread of the infection. The validation of an instrument that can easily identify a suspected COVID-19 case by attributing a score to each symptom related to COVID-19 can be of great importance in facilitating the containment of the epidemic. Our proposed cutoff score seems worthy of validation for use in broader populations to confirm its clinimetric properties. In the event of its validation, our cutoff score might be useful in selecting people who require serological and molecular diagnostic tests for COVID-19.

The availability and accessibility of diagnostic tests for the SARS-CoV-2 coronavirus have proven to be key in containing the COVID-19 pandemic. The early identification of subjects who test positive for COVID-19 (ie, via molecular and serological tests) among people with specific symptoms or people who are at risk is crucial for limiting the spread of the infection. The tool we validated responds to the need for readily identifying a suspected COVID-19 case, by attributing a score to each symptom related to COVID-19. Although our validation was satisfactory, our proposed cutoff score seems worthy of further testing in larger populations in order to confirm its clinimetric properties and usefulness in selecting people who require serological and molecular diagnostic tests for COVID-19.

Although the EPICOVID19 DS tool can be used as a public health prevention instrument, directing subjects to a self-assessment tool without warning may trigger panic, alarm, and concern among the screened population. Furthermore, the limitations of our study must be considered. First, participation in this study was voluntary, and the sample was not representative of the general population. This means that potential selection biases must be taken into consideration. Second, data were collected from a highly educated, young-adult convenience population sample with low multimorbidity. This was a result of the phase I EPICOVID19 study [10], and such a sample is expected in studies that involve a web-based questionnaire that is promoted via email invitation. Third, in the context of a pandemic, our survey might have interested people who had no opportunity to report symptoms to clinicians. Moreover, the effect of recall bias cannot be excluded among the participants who tested positive for COVID-19 or presented with symptoms related to SARS-COV2 infection. The fourth limitation of our study is the small sample size in the analysis of the 2 age groups (ie, subjects aged <60 and ≥60 years). Given these limitations, the adoption of the EPICOVID19 DS should be considered with caution. The procedures outlined for the development of the EPICOVID19 DS can be applied iteratively as new data is collected, to continue the refinement of this potentially valuable clinical decision support tool.

In conclusion, the proposed EPICOVID19 DS seems worthy of further testing in different scenarios and populations to achieve a comprehensive understanding of its clinimetric properties for both low-prevalence and high-prevalence COVID-19 settings, and its aptitude for capturing disease severity data. This will allow us to define the boundaries of its use and identify optimal indicators to assist clinicians with the early recognition of COVID-19.

Acknowledgments

We would like to thank Professor Mario Grassi of the University of Pavia for his suggestions regarding statistical analysis and Dr Michael N Liebman for his support in the final revision of the manuscript contents. We would also like to thank them both for their English language support. The authors would like to thank all the participants who took part in this study and made this study possible.

Authors' Contributions

SM, LB, FA, and FP were responsible for the study concept and design. LB, LF, and SP were responsible for the literature search. FA, LF, and SP were responsible for the acquisition of data. LB, FA, FB, FP, and SM were responsible for the analysis and interpretation of data. LB, LF, SM, SP, and FB were responsible for drafting the manuscript. MG, AG, RAI, CP, MA, and GP were responsible for critically revising the manuscript for important intellectual content. LB and SM were responsible for statistical analysis. GP, SM, CT, MN, NJ, AS, CP, and MA critically revised the manuscript for important intellectual content. All authors participated in data interpretation. All authors read and approved the final version of the manuscript. The corresponding author, SM, attests that all listed authors meet authorship criteria and that no others who meet the criteria have been omitted.

Conflicts of Interest

AG received consultancy fees from Mylan and educational support from Gilead.

Multimedia Appendix 1
Phase I EPICOVID19 questionnaire, Phase 2 Valid Symptoms Section of the EPICOVID19 questionnaire, and statistical appendix.
[DOCX File , 79 KB - jmir_v23i1e23897_app1.docx]

References

1. Coronavirus (COVID-19) events as they happen. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2020-12-04]
2. COVID-19 map. Johns Hopkins Coronavirus Resource Centre. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-07-24]
3. Giacomelli A, Ridolfo AL, Milazzo L, Oreni L, Bernacchia D, Siano M, et al. 30-day mortality in patients hospitalized with COVID-19 during the first wave of the Italian epidemic: A prospective cohort study. Pharmacol Res 2020 Aug;158:104931 [FREE Full text] [doi: 10.1016/j.phrs.2020.104931] [Medline: 32446978]
4. COVID-19 Clinical Research Coalition. Global coalition to accelerate COVID-19 clinical research in resource-limited settings. Lancet 2020 Apr 25;395(10233):1322-1325 [FREE Full text] [doi: 10.1016/S0140-6736(20)30798-4] [Medline: 32247324]
5. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 - Studies Needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/NEJMp2002125] [Medline: 32074416]
6. 6 W. Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, 17 March 2020. World Health Organization. URL: https://apps.who.int/iris/handle/10665/331656 [accessed 2020-06-24]
7. Green K, Graziadio S, Turner P, Fanshawe T, Allen J. Molecular and antibody point-of-care tests to support the screening, diagnosis and monitoring of COVID-19. The Centre for Evidence-Based Medicine. 2020 Apr 07. URL: https://www.

cebm.net/covid-19/molecular-and-antibody-point-of-care-tests-to-support-the-screening-diagnosis-and-monitoring-of-covid-19/ [accessed 2020-12-24]

8. Global surveillance for COVID-19 caused by human infection with COVID-19 virus: interim guidance, 20 March 2020. World Health Organization. URL: https://apps.who.int/iris/handle/10665/331506 [accessed 2020-12-24]
9. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
10. Adorni F, Prinelli F, Bianchi F, Giacomelli A, Pagani G, Bernacchia D, et al. Self-Reported Symptoms of SARS-CoV-2 Infection in a Nonhospitalized Population in Italy: Cross-Sectional Study of the EPICOVID19 Web-Based Survey. JMIR Public Health Surveill 2020;6(3) [FREE Full text] [doi: 10.2196/21866] [Medline: 32650305]
11. General Data Protection Regulation (GDPR) — Official Legal Text. General Data Protection Regulation. URL: https://gdpr-info.eu/ [accessed 2020-09-24]
12. Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med 2020 Jun;26(6):845-848. [doi: 10.1038/s41591-020-0897-1] [Medline: 32350462]
13. Chan JF, Yuan S, Kok K, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020 Feb 15;395(10223):514-523 [FREE Full text] [doi: 10.1016/S0140-6736(20)30154-9] [Medline: 31986261]
14. Giacomelli A, Pezzati L, Conti F, Bernacchia D, Siano M, Oreni L, et al. Self-reported Olfactory and Taste Disorders in Patients With Severe Acute Respiratory Coronavirus 2 Infection: A Cross-sectional Study. Clin Infect Dis 2020 Jul 28;71(15):889-890 [FREE Full text] [doi: 10.1093/cid/ciaa330] [Medline: 32215618]
15. Pallanti S. Importance of SARs-Cov-2 anosmia: From phenomenology to neurobiology. Compr Psychiatry 2020 Jul;100:152184 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152184] [Medline: 32422426]
16. Menni C, Sudre CH, Steves CJ, Ourselin S, Spector TD. Quantifying additional COVID-19 symptoms will save lives. Lancet 2020 Jun 20;395(10241):e107-e108 [FREE Full text] [doi: 10.1016/S0140-6736(20)31281-2] [Medline: 32505221]
17. Cholankeril G, Podboy A, Aivaliotis VI, Tarlow B, Pham EA, Spencer SP, et al. High Prevalence of Concurrent Gastrointestinal Manifestations in Patients With Severe Acute Respiratory Syndrome Coronavirus 2: Early Experience From California. Gastroenterology 2020 Aug;159(2):775-777 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.008] [Medline: 32283101]
18. Sarma P, Kaur H, Kaur H, Bhattacharyya J, Prajapat M, Shekhar N, et al. Ocular Manifestations and Tear or Conjunctival Swab PCR Positivity for 2019-nCoV in Patients with COVID-19: A Systematic Review and Meta-Analysis. SSRN. Preprint posted online on April 15, 2020. [doi: 10.2139/ssrn.3566161]
19. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
20. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
21. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. J Med Internet Res 2020 Jun 30;22(6):e19636 [FREE Full text] [doi: 10.2196/19636] [Medline: 32544071]
22. Bénézit F, Le Turnier P, Declerck C, Paillé C, Revest M, Dubée V, RAN COVID Study Group. Utility of hyposmia and hypogeusia for the diagnosis of COVID-19. Lancet Infect Dis 2020 Sep;20(9):1014-1015 [FREE Full text] [doi: 10.1016/S1473-3099(20)30297-8] [Medline: 32304632]
23. Sun P, Qie S, Liu Z, Ren J, Li K, Xi J. Clinical characteristics of hospitalized patients with SARS-CoV-2 infection: A single arm meta-analysis. J Med Virol 2020 Jun;92(6):612-617 [FREE Full text] [doi: 10.1002/jmv.25735] [Medline: 32108351]
24. Jain V, Yuan JM. Systematic review and meta-analysis of predictive symptoms and comorbidities for severe COVID-19 infection. MedRxiv. Preprint posted online on March 16, 2020. [doi: 10.1101/2020.03.15.20035360]
25. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
26. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020 May 12;323(18):1775-1776. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]
27. Preventing and managing COVID-19 across long-term care services: policy brief, 24 July 2020. World Health Organization. URL: https://apps.who.int/iris/handle/10665/333074 [accessed 2020-12-24]
28. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA 2020 May 12;323(18):1843-1844 [FREE Full text] [doi: 10.1001/jama.2020.3786] [Medline: 32159775]

29. Yong SEF, Anderson DE, Wei WE, Pang J, Chia WN, Tan CW, et al. Connecting clusters of COVID-19: an epidemiological and serological investigation. Lancet Infect Dis 2020 Jul;20(7):809-815 [FREE Full text] [doi: 10.1016/S1473-3099(20)30273-5] [Medline: 32330439]
30. Manski CF. Bounding the Predictive Values of COVID-19 Antibody Tests. MedRxiv. Posted online on May 18, 2020 2020. [doi: 10.3386/w27226]
31. Mathur G, Mathur S. Antibody Testing for COVID-19: Can It Be Used as a Screening Tool in Areas With Low Prevalence? Am J Clin Pathol 2020 Jun 08;154(1):1-3 [FREE Full text] [doi: 10.1093/ajcp/aqaa082] [Medline: 32412044]
32. Razai MS, Doerholt K, Ladhani S, Oakeshott P. Coronavirus disease 2019 (covid-19): a guide for UK GPs. BMJ 2020 Mar 05;368:m800. [doi: 10.1136/bmj.m800] [Medline: 32144127]
33. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]

Abbreviations

AUC: area under the curve
CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
EPICOVID19 DS: EPICOVID19 diagnostic scale
HOMALS: homogeneity analysis by means of alternating least squares
IgG: immunoglobulin G
IgM: immunoglobulin M
NPS: nasopharyngeal swab
NPV: negative predictive value
PHP: hypertext preprocessor
PPV: positive predictive value
RMSEA: root mean square error of approximation
SRSR: standardized root mean square residual
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 27.08.20; peer-reviewed by S Syed, C Miranda, A Azzam; comments to author 09.09.20; revised version received 28.09.20; accepted 29.09.20; published 06.01.21.

Please cite as:
Bastiani L, Fortunato L, Pieroni S, Bianchi F, Adorni F, Prinelli F, Giacomelli A, Pagani G, Maggi S, Trevisan C, Noale M, Jesuthasan N, Sojic A, Pettenati C, Andreoni M, Antonelli Incalzi R, Galli M, Molinaro S
Rapid COVID-19 Screening Based on Self-Reported Symptoms: Psychometric Assessment and Validation of the EPICOVID19 Short Diagnostic Scale
J Med Internet Res 2021;23(1):e23897
URL: http://www.jmir.org/2021/1/e23897/
doi:10.2196/23897
PMID:33320825

Original Paper

Accurately Differentiating Between Patients With COVID-19, Patients With Other Viral Infections, and Healthy Individuals: Multimodal Late Fusion Learning Approach

Ming Xu[1,2,3*], MD, PhD; Liu Ouyang[4*], MD; Lei Han[1,2*], MD, PhD; Kai Sun[5*], MD; Tingting Yu[6*], MD; Qian Li[7*], MD; Hua Tian[8*], MSc; Lida Safarnejad[9,10], MSc, PhD; Hengdong Zhang[1,2], MD, PhD; Yue Gao[1], MSc; Forrest Sheng Bao[11], PhD; Yuanfang Chen[12], MSc; Patrick Robinson[3], MD, FACP, FIDSA; Yaorong Ge[10], PhD; Baoli Zhu[2], MSc; Jie Liu[13], MD, PhD; Shi Chen[3,14], PhD

[1]Department of Occupational Disease Prevention, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

[2]Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

[3]Department of Public Health Sciences, College of Health and Human Services, University of North Carolina Charlotte, Charlotte, NC, United States

[4]Department of Orthopedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[5]Department of Emergency Medicine, The First Hospital of Nanjing Medical University, Nanjing, China

[6]Department of Medical Genetics, School of Basic Medical Science Jiangsu Key Laboratory of Xenotransplantation, Nanjing Medical University, Nanjing, China

[7]Department of Pediatrics, Affiliated Kunshan Hospital of Jiangsu University, Kunshan, China

[8]Department of Acute Infectious Disease Control and Prevention, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

[9]School of Medicine, Stanford University, Stanford, CA, United States

[10]Department of Software and Information Systems, College of Computing and Informatics, University of North Carolina Charlotte, Charlotte, NC, United States

[11]Department of Computer Science, Iowa State University, Ames, IA, United States

[12]Institute of HIV/AIDS/STI Prevention and Control, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China

[13]Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[14]School of Data Science, University of North Carolina Charlotte, Charlotte, NC, United States

*these authors contributed equally

Corresponding Author:
Jie Liu, MD, PhD
Department of Radiology
Union Hospital, Tongji Medical College
Huazhong University of Science and Technology
1277 Jiefang Avenue
Wuhan
China
Phone: 86 13469981699
Email: liu_jie0823@163.com

Abstract

Background: Effectively identifying patients with COVID-19 using nonpolymerase chain reaction biomedical data is critical for achieving optimal clinical outcomes. Currently, there is a lack of comprehensive understanding in various biomedical features and appropriate analytical approaches for enabling the early detection and effective diagnosis of patients with COVID-19.

Objective: We aimed to combine low-dimensional clinical and lab testing data, as well as high-dimensional computed tomography (CT) imaging data, to accurately differentiate between healthy individuals, patients with COVID-19, and patients with non-COVID viral pneumonia, especially at the early stage of infection.

Methods: In this study, we recruited 214 patients with nonsevere COVID-19, 148 patients with severe COVID-19, 198 noninfected healthy participants, and 129 patients with non-COVID viral pneumonia. The participants' clinical information (ie, 23 features), lab testing results (ie, 10 features), and CT scans upon admission were acquired and used as 3 input feature modalities. To enable the late fusion of multimodal features, we constructed a deep learning model to extract a 10-feature high-level representation of

CT scans. We then developed 3 machine learning models (ie, k-nearest neighbor, random forest, and support vector machine models) based on the combined 43 features from all 3 modalities to differentiate between the following 4 classes: nonsevere, severe, healthy, and viral pneumonia.

Results: Multimodal features provided substantial performance gain from the use of any single feature modality. All 3 machine learning models had high overall prediction accuracy (95.4%-97.7%) and high class-specific prediction accuracy (90.6%-99.9%).

Conclusions: Compared to the existing binary classification benchmarks that are often focused on single-feature modality, this study's hybrid deep learning-machine learning framework provided a novel and effective breakthrough for clinical applications. Our findings, which come from a relatively large sample size, and analytical workflow will supplement and assist with clinical decision support for current COVID-19 diagnostic methods and other clinical applications with high-dimensional multimodal biomedical features.

(J Med Internet Res 2021;23(1):e25535) doi:10.2196/25535

KEYWORDS

COVID-19; machine learning; deep learning; multimodal; feature fusion; biomedical imaging; diagnosis support; diagnosis; imaging; differentiation; testing; diagnostic

Introduction

COVID-19 is an emerging major biomedical challenge for the entire health care system [1]. Compared to severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), COVID-19 has much higher infectivity. COVID-19 has also spread much faster across the globe than other coronavirus diseases. Although COVID-19 has a relatively lower case fatality rate than SARS and MERS, the overwhelmingly large number of diagnosed COVID-19 cases, as well as the many more undiagnosed COVID-19 cases, has endangered health care systems and vulnerable populations during the COVID-19 pandemic. Therefore, the early and accurate detection and intervention of COVID-19 are key in effectively treating patients, protecting vulnerable populations, and containing the pandemic at large.

Currently, the gold standard for the confirmatory diagnosis of COVID-19 is based on molecular quantitative real-time polymerase chain reaction (qRT-PCR) and antigen testing for the disease-causing SARS-CoV-2 virus [2-4]. Although these tests are the gold standard for COVID-19 diagnosis, they suffer from various practical issues, including reliability, resource adequacy, reporting lag, and testing capacity across time and space [5]. To help frontline clinicians diagnose COVID-19 more effectively and efficiently, other diagnostic methods have also been explored and used, including medical imaging (eg, X-ray scans and computed tomography [CT] scans [6]), lab testing (eg, various blood biochemistry analyses [7-10]), and identifying common clinical symptoms [11]. However, these methods do not directly detect the disease-causing SARS-CoV-2 virus or the SARS-CoV-2 antigen. Therefore, these methods do not have the same conclusive power that confirmatory molecular diagnostic methods have. Nevertheless, these alternative methods help clinicians with inadequate resources detect COVID-19, differentiate patients with COVID-19 from patients without COVID-19 and noninfected individuals, and triage patients to optimize health care system resources [12,13]. When applied appropriately, these supplementary methods, which are based on alternative biomedical evidence, can help mitigate the COVID-19 pandemic by accurately identifying patients with COVID-19 as early as possible.

Currently, CT scans can be analyzed to differentiate patients with COVID-19, especially those in a severe clinical state, from healthy people or patients with non-COVID infections. Patients with COVID-19 usually present the typical ground-glass opacity (GGO) characteristic on CT images of the thoracic region. A recent study has reported a 98% COVID-19 detection rate based on a 51-patient sample without a comparison group [14]. Detection rates that ranged between 60% and 93% were also reported in another study on 1014 participants with a comparison group [15]. Furthermore, the recent advances in data-driven deep learning (DL) methods, such as convolutional neural networks (CNNs), have demonstrated the ability to detect COVID-19 in patients. On February 2020, Hubei, China adopted CT scans as the official clinical COVID-19 diagnostic method in addition to molecular confirmatory diagnostic methods for COVID-19, in accordance with the nation's diagnosis and treatment guidance [2]. However, the effectiveness of using DL methods to further differentiate SARS-CoV-2 infection from clinically similar non-COVID viral infections still needs to be explored and evaluated.

With regard to places where molecular confirmatory diagnoses are not immediately available, symptoms are often used for quickly evaluating presumed patients' conditions and supporting triage [13,16,17]. Checklists have been developed for self-evaluating the risk of developing COVID-19. These checklists are based on clinical information, including symptoms, preexisting comorbidities, and various demographic, behavioral, and epidemiological factors. However, these clinical data are generally used for qualitative purposes (eg, initial assessment) by both the public and clinicians [18]. Their effectiveness in providing accurate diagnostic decision support is largely underexplored and unknown.

In addition to biomedical imaging and clinical information, recent studies on COVID-19 have shown that laboratory testing, such as various blood biochemistry analyses, is also a feasible method for detecting COVID-19 in patients, with reasonably high accuracy [19,20]. The rationale is that the human body is a unity. When people are infected with SARS-CoV-2, the clinical consequences can be observed not only from apparent symptoms, but also from hematological biochemistry changes. Due to the challenge of asymptomatic SARS-CoV-2 infection, other types of biomedical information, such as lab testing results, can be used to provide alternative and complementary diagnostic decision support evidence. It is possible that our current definition and understanding of asymptomatic infection can be extended with more intrinsic, quantitative, and subtle medical features, such as blood biochemistry characteristics [21,22].

Despite the tremendous advances in obtaining alternative and complementary diagnostic evidence for COVID-19 (eg, CT scans, chest X-rays, clinical information, and various blood biochemistry characteristics), there are still substantial clinical knowledge gaps and technical challenges that hinder our efforts in harnessing the power of various biomedical data. First, most recent studies have usually focused on one of the multiple modalities of diagnostic data, and these studies have not considered the potential interactions between and added interpretability of these modalities. For example, can we use both CT scan and clinical information to develop a more accurate COVID-19 decision support system [23]? As stated earlier, the human body acts as a unity against SARS-CoV-2 infection. Biomedical imaging and clinical approaches can be used to evaluate different aspects of the clinical consequences of COVID-19. By combining the different modalities of biomedical information, a more comprehensive characterization of COVID-19 can be achieved. This is referred to as multimodal biomedical information research.

Second, while there are ample accurate DL algorithms/models/tools, especially in biomedical imaging, most of them focus on differentiating patients with COVID-19 from noninfected healthy individuals. A moderately trained radiologist can differentiate CT scans of patients with COVID-19 from those of healthy individuals with high accuracy, making current efforts in developing supplicated DL algorithms not clinically useful for solving the binary classification problem [14]. The more critical and urgent clinical issue is not only being able to differentiate patients with COVID-19 from noninfected healthy individuals, but also being able to differentiate SARS-CoV-2 infection from non-COVID viral infections [24,25]. Patients with non-COVID viral infection present with GGO in their CT scans of the thoracic region as well. Therefore, the specificity of GGO as a diagnostic criterion of COVID-19 is low [15]. In addition, patients with nonsevere COVID-19 and patients with non-COVID viral infection share several common symptoms, which are easy to confuse [26]. Therefore, for frontline clinicians, effectively differentiating nonsevere COVID-19 from non-COVID viral infection is a challenging task without readily available and reliable confirmatory molecular tests at admission. Incorrectly diagnosing severe COVID-19 as nonsevere COVID-19 may result in missing the critical window of intervention. Similarly, differentiating asymptomatic and presymptomatic patients, including those with nonsevere COVID-19, from noninfected healthy individuals is another major clinical challenge [27]. Incorrectly diagnosing patients without COVID-19 or healthy individuals and treating them alongside patients with COVID-19 will substantially increase their risk of exposure to the virus and result in health care–associated infections. There is an urgent need for a multinomial classification system that can detect patients with COVID-19, including patients with asymptomatic COVID-19, patients with non-COVID viral infection, and healthy individuals, all at once, rather than a system that analyzes several independent binary classifiers in parallel [28].

The third major challenge addresses the computational aspect of harnessing the power of various biomedical data. Due to the novelty of the COVID-19 pandemic, human clinicians have varying degrees of understanding and experience with regard to COVID-19, which can lead to inconsistencies in clinical decision making. Harnessing the power of multimodal biomedical information from combined imaging, clinical, and lab testing data can be the basis of a more objective, data-driven, analytical framework. In theory, such a framework can provide a more comprehensive understanding of COVID-19 and a more accurate decision support system that can differentiate between patients with severe or nonsevere COVID-19, patients with non-COVID viral infection, and healthy individuals all at once. However, biomedical imaging data, such as CT data, with a high-dimensional feature space do not integrate well with low-dimensional clinical and lab testing data. Current studies have usually only described the association between biomedical imaging and clinical features [15,29-33], and the potential power of an accurate decision support tool has not been reported. Technically, CT scans are usually processed with DL methods, including the CNN method, independently from other types of biomedical data processing methods. Low-dimensional clinical and lab testing data are usually analyzed with traditional hypothesis-driven methods (eg, binary logistic regression or multinomial classification) or other non-DL machine learning (ML) methods, such as the random forest (RF), support vector machine (SVM), and k-nearest neighbor (kNN) methods. The huge discrepancy of feature space dimensionality between CT scan and clinical/lab testing data makes multimodal fusion (ie, the direct combination of the different aspects of biomedical information) especially challenging [34].

To fill these knowledge gaps and overcome the technical challenge of effectively analyzing multimodal biomedical information, we propose the following study objective: we aimed to clinically and accurately differentiate between patients with nonsevere COVID-19, patients with severe COVID-19, patients with non-COVID viral pneumonia, and healthy individuals all at once. To successfully fulfill this much-demanded clinical objective, we developed a novel hybrid DL-ML framework that harnesses the power of a wide array of complex multimodal data via feature late fusion. The clinical objective and technical approach of this study synergistically complement each other to form the basis of an accurate COVID-19 diagnostic decision support system.

Methods

Participant Recruitment

We recruited a total of 362 patients with confirmed COVID-19 from Wuhan Union Hospital between January 2020 and March 2020 in Wuhan, Hubei Province, China. COVID-19 was confirmed based on 2 independent qRT-PCR tests. For this study, we did not aggregate patients with COVID-19 under the same class because the clinical characteristics of nonsevere and severe COVID-19 were distinct. Patients' COVID-19 status was confirmed upon admission. The recruited patients were further categorized as being in severe (n=148) or nonsevere (n=214) clinical states based on their prognosis at 7-14 days after initial admission. This step ensured the development of an early detection system for when the initial conditions of patients with COVID-19 were not severe upon admission. Patients in the severe state group were identified by having 1 of the following 3 clinical features: (1) respiratory rate>30 breaths per minute, (2) oxygen saturation<93% at rest, and (3) arterial oxygen partial pressure/fraction of inspired oxygen<300 mmHg (ie, 40 kPa). These clinical features are based on the official COVID-19 Diagnosis and Treatment Plan from the National Health Commission of China [2], as well as guidelines from the American Thoracic Society [35]. The noninfected group included 198 healthy individuals without any infections. These participants were from the 2019 Hubei Provincial Centers for Disease Control and Prevention regular annual physical examination cohort. This group represented a baseline healthy group, and they were mainly used as a comparison group for patients with nonsevere COVID-19, especially those who presented with inconspicuous clinical symptoms.

In order to differentiate patients with COVID-19, especially those with nonsevere COVID-19, from patients with clinically similar non-COVID viral infection, we also included another group of 129 patients diagnosed with non-COVID viral pneumonia in this study. It should be noted that the term “viral pneumonia” was an umbrella term that included diseases caused by more than 1 type of virus, such as the influenza virus and adenovirus. However, in clinical practice, it would be adequate to detect and differentiate between SARS-CoV-2 infection and non-COVID viral infections for initial triaging. Therefore, we recruited 129 participants with confirmed non-COVID viral infection from Kunshan Hospital, Suzhou, China. The reality was that most health care resources were optimized for COVID-19, and some patients who presented with COVID-19–like symptoms or GGOs were clinically diagnosed with COVID-19 without the use of confirmatory qRT-PCR tests in Hubei, especially during February 2020. Therefore, it was not possible to recruit participants with non-COVID viral infection in Hubei during the same period that we recruited patients with COVID-19.

In summary, the entire study sample was comprised of the following 4 mutually exclusive multinomial participant classes: severe COVID-19 (n=148), nonsevere COVID-19 (n=214), non-COVID viral infection (n=129), and noninfected healthy (n=198). This study was conducted in full compliance with the Declaration of Helsinki. This study was rigorously evaluated and approved by the institutional review board committees of Jiangsu Provincial Center for Disease Control and Prevention (approval number JSJK2020-8003-01). All participants were comprehensively told about the details of the study. All participants signed a written informed consent form before being admitted.

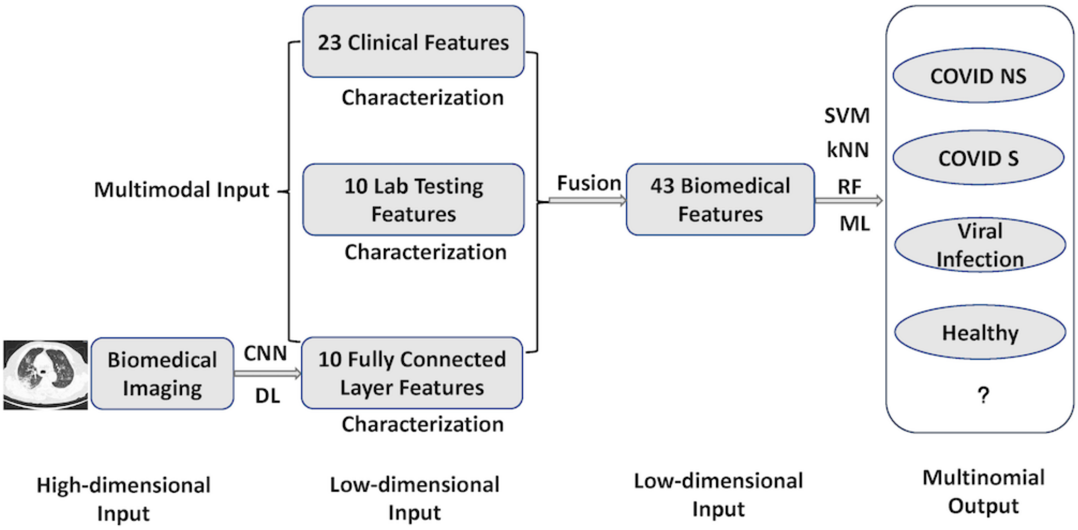
Medical Feature Selection and Description

Patient participants, including those in the severe COVID-19, nonsevere COVID-19, and non-COVID viral infection classes, were screened upon initial admission into hospitals. Their clinical information, including preexisting comorbidities, symptoms, demographic characteristics, epidemiological characteristics, and other clinical data, were recorded. For the noninfected healthy class, participants’ clinical data were extracted from the Hubei Provincial Centers for Disease Control and Prevention physical examination record system. Patient-level sensitive information, including name and exact residency, were completely deidentified. After comparing the different classes, the following 23 clinical features were selected for this study: smoking history, hypertension, type-2 diabetes, cardiovascular disease (ie, any type), chronic obstructive pulmonary disease, fever, low fever, medium fever, high fever, sore throat, coughing, phlegm production, headache, feeling chill, muscle ache, feelings of fatigue, chest congestion, diarrhea, loss of appetite, vomiting, old age (ie, >50 years; dichotomized and encoded as old), and gender. These clinical data were dichotomized as either having the condition (score=1) or not having the condition (score=0) (Figure 1). It should be noted that several clinical features, especially symptoms, were self-reported by the patients. A more comprehensive definition and description of clinical features are provided in Multimedia Appendix 1. The prevalence (ie, the number of participants that have a given feature over the total number of participants in the class) of each clinical feature was computed across the 4 classes. For the 0-1 binary clinical features, a pairwise z-test was applied to detect any substantial differences in the prevalence (ie, proportion) of these features between classes.

The lab testing features were extracted from participants’ electronic health records (Figure 1). Only the features from lab tests that were performed at the time of admission were included. Noninfected healthy participants’ blood samples were taken during their annual physical examination. We selected lab testing features that were present in at least 90% of participants in any of the 4 classes (ie, severe COVID-19, nonsevere COVID-19, non-COVID viral infection, and noninfected healthy). After screening, the following 10 features were included: white blood cell count, hemoglobin level, platelet count, neutrophil count, neutrophil percent, lymphocyte count, lymphocyte percent, C-reactive protein level, total bilirubin level, and creatine level. Features in the lab testing modality all had continuous numeric values, which were different from the 0-1 binary values in the clinical feature modality. The distributions of these lab testing features were compared across the 4 classes by using a 2-sided Kolmogorov-Smirnov test. In addition, we also applied the Kruskal-Wallis test for multiple comparisons across the 4 classes for the top 3 most differentiating features, which were identified later by an ML workflow. The Kolmogorov-Smirnov test was applied during initial screening to investigate whether the values of the same biomedical feature were distributed differently between 2 classes. The nonparametric Kruskal-Wallis test was chosen because it could rigorously compare classes and provide robust results for nonnormal data. The test was able to accommodate more than 2 classes (ie, multinomial classes) in this study.

Each participant underwent CT scans of the thoracic region in the radiology department. Toshiba Activion 16 multislice CT scanners were used to perform CT scanning at around 120 kVp with a tube current of 50 mA. We obtained 50 CT images per scan, and each image had the following characteristics: slice thickness=2 mm, voxel size=1.5 mm, and image resolution=512×512 pixels. Each participant underwent an average of 50 CT scans. The total number of CT images obtained in this study was over 30,000. CT images were archived and presented as DICOM (Digital Imaging and Communications in Medicine) images for DL.

Figure 1. Multimodal feature late fusion and multinomial classification workflow. A deep learning convolutional neural network was applied to computed tomography images for representation learning and extracting 10 features from a customized fully connected layer. These 10 features were merged with other modality data through feature late fusion. In the machine learning stage of the workflow, each of the 3 machine learning models (ie, the support vector machine, k-nearest neighbor, and random forest models) worked independently to provide their respective outputs. kNN: k-nearest neighbor; ML: machine learning; RF: random forest; SVM; support vector machine.



The Multinomial Classification Objective

The main research goal of this study was to accurately differentiate between patients with severe COVID-19, patients with nonsevere COVID-19, patients with non-COVID viral infection, and noninfected healthy individuals from a total of N participants all at once. Therefore, a formula was developed to address the multinomial output classification problem. The following equation uses 1 of the 4 mutually exclusive output classes (ie, H=noninfected healthy, V=non-COVID viral pneumonia, NS=nonsevere COVID-19, and S=severe COVID-19) of an individual (ie, i), as follows:

$$f(X_c,X_l,X_m)_i = \{H,V,NS,S\},\ i = 1...N \quad \textbf{(1)}$$

In this equation, the inputs were individuals' (ie, i) multimodal features of binary clinical information (ie, X_c), continuous lab test results (ie, X_l), and CT imaging (ie, X_m). The major advantage of our study was that we were able to classify 4 classes all at once, instead of developing several binary classifiers in parallel.

The Hybrid DL-ML Approach: Feature Late Fusion

As stated earlier, the voxel level of CT imaging data does not integrate well with low-dimensional clinical and lab testing features. In this study, we proposed a feature late fusion approach via the use of hybrid DL and ML models. Technically, DL is a type of ML that uses deep neural networks (eg, CNNs are a type of deep neural network). In this study, we colloquially used the term “machine learning” to refer to more traditional, non-DL types of ML (eg, RF ML), in contrast with DL that focuses on deep neural networks. An important consideration in the successful late fusion of multimodality features is the representation learning of the high-dimensional CT features.

For each CT scan of each participant, we constructed a customized residual neural network (ResNet) [36-39], which is a specific architecture for DL CNNs. A ResNet is considered a mature CNN architecture with relatively high performance across different tasks. Although other CNN architectures exist (eg, EfficientNet, VGG-16, etc), the focus of this study was not to compare different architectures. Instead, we wanted to deliver the best performance possible with a commonly used CNN architecture (ie, ResNet) for image analysis.

By constructing a ResNet, we were able to transform the voxel-level imaging data into a high-level representation with significantly fewer features. After several convolution and max pooling layers, the ResNet reached a fully connected (FC; ie, FC1 layer) layer before the final output layer, thereby enabling the delivery of actual classifications. In the commonly used ResNet architecture, the FC layer is a 1×512 vector, which is relatively closer in dimensionality to clinical information (ie, 1×23 vector) and lab testing (ie, 1×10 vector) feature modalities. However, the original FC layer from the ResNet was still much larger than the other 2 modalities. Therefore, we added another

FC layer (ie, FC2 layer) after the FC1 layer, but before the final output layer. In this study, the FC2 layer was set to have a 1×10 vector dimension (ie, 10 elements in the vector) to match the dimensionality of the other 2 feature modalities. Computationally, the FC2 layer served as the low-dimensional, high-level representation of the original CT scan data. The distributions of the 10 features extracted from the ResNet in the FC2 layer were compared across the 4 classes with the Kolmogorov-Smirnov test. The technical details of this customized ResNet architecture are provided in Multimedia Appendix 2.

Once low-dimensional high-level features were extracted from CT data via the ResNet CNN, we performed multimodal feature fusion. The clinical information, lab testing, and FC2 layer features of each participant (ie, i) were combined into a single 1× 43 (ie, 1×[23+10+10]) row vector. The true values of the output were the true observed classes of the participants. Technically, the model would try to predict the outcome as accurately as it could, based on the observed classes.

The Hybrid DL-ML Approach: Modeling

After deriving the feature matrix, we applied ML models for the multinomial classification task. In this study, 3 different types of commonly used ML models were considered, as follows: the RF, SVM, and kNN models. An RF model is a decision-tree–based ML model, and the number of tree hyperparameters was set at 10, which is a relatively small number compared to the number of input features needed to avoid potential model overfitting. Other RF hyperparameters in this study included the Gini impurity score to determine tree split, at least 2 samples to split an internal tree, and at least 1 sample at a leaf node. All default hyperparameter settings, including those of the SVM and RF models, were based on the scikit-learn library in Python. An SVM model is a model of maximum hyperplane and L-2 penalty; radial basis function kernels and a gamma value of 1/43 (ie, the inverse of the total number of features) were used as hyperparameter values in this study. kNN is a nonparametric instance-based model; the following hyperparameter values were used in this study: k=5, uniform weights, tree leaf size=30, and p=2. These 3 models are technically distinct types of ML models. We aimed to investigate whether specific types of ML models and multimodal feature fusion would contribute to developing an accurate COVID-19 classifier for clinical decision support.

We evaluated each respective ML model with 100 independent runs. Each run used a different randomly selected dataset comprised of 80% of the original data for training, and the remaining 20% of data were used to test and validate the model. Performing multiple runs instead of a single run revealed how robust the model was, despite system stochasticity. The 80%-20% split of the original data for separate training and testing sets also ensured that potential model overfitting and increased model generalizability could be avoided. In addition, RF models use bagging for internal validation based on out-of-bag errors (ie, how the "tree" would split out in the "forest" model).

After each run, important ML performance metrics, including accuracy, sensitivity, precision, and F1 score, were computed for the test set. We reported the overall performance of the ML models first. These different metrics evaluated ML models based on different aspects. In this study, we also considered 3 different approaches for calculating the overall performance of multinomial outputs, as follows: a micro approach (ie, the one-vs-all approach), a macro approach (ie, unweighted averages; each of the 4 classes were given the same 25% weights), and a weighted average approach based on the percentage of each class in the entire sample.

In addition, because the output in this study was multinomial instead of binary, each class had its own performance metrics. We aggregated these performance metrics across 100 independent runs, determined each metric's distribution, and evaluated model robustness based on these distributions. If ML performance metrics in the testing set had a small variation (ie, small standard errors), then the model was considered robust against model input changes, thereby allowing it to reveal the intrinsic pattern of the data. This was because in each run, a different randomly selected dataset (ie, 80% of the original data) was selected to train the model.

An advantage that the RF model had over SVM and kNN models was that it had relatively clearer interpretability, especially when interpreting feature importance. After developing the RF model based on the training set, we were able to rank the importance of input features based on their corresponding Gini impurity score from the RF model [40,41]. It should be noted that only the training set was used to compute Gini impurity, not the test set. We then assessed the top contributing features' clinical relevance to COVID-19.

We also developed and evaluated the performance of single-modality (ie, using clinical information, lab testing, and CT features individually) ML models. The performance results were used as baseline conditions. The models' performance results were then compared to the multimodal classifications to demonstrate the potential performance gain of the feature fusion of different feature modalities. In this study, each individual ML model (ie, the RF, SVM, and kNN models) was independently evaluated, and the respective results were reported, without combining the prediction of the final output class.

The deep learning CNN and late fusion machine learning codes were developed in Python with various supporting packages, such as scikit-learn.

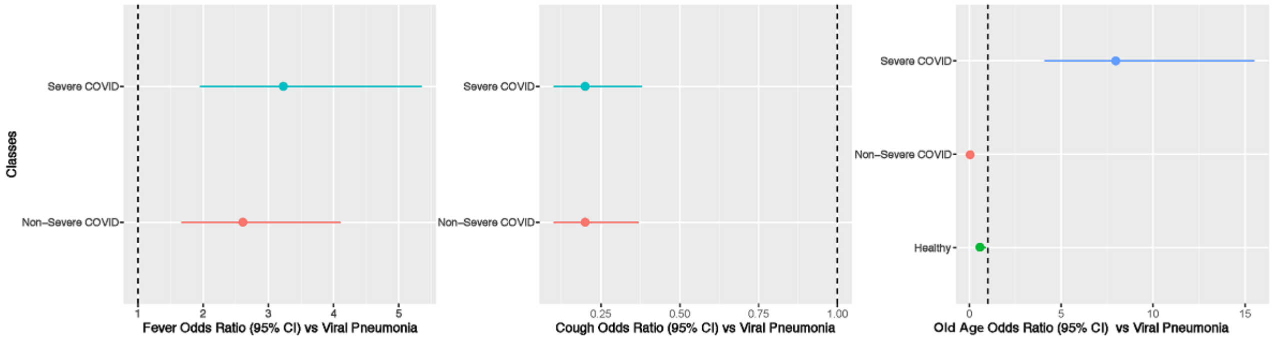
Results

Clinical Characterization of the 4 Classes

Detailed demographic, clinical, and lab testing results among four classes were provided in supplementary Table S1. We compared clinical features across the 4 classes. The prevalence of each feature in all 4 classes is shown in Multimedia Appendix 3. In general, most clinical features varied substantially between the nonsevere COVID-19, severe COVID-19, and non-COVID viral pneumonia classes. It should be noted that all symptom feature values, except gender and age group (ie, >50 years) values, in the noninfected healthy class were set to 0, so that they could be used as a reference. Based on the 2-sample z-test of proportions, the nonsevere COVID-19 and severe COVID-19 classes differed significantly (*P*<.05) in 10 out of 22 symptom features, including comorbidities such as hypertension (*P*<.001), diabetes (*P*<.001), cardiovascular diseases (*P*<.001), and chronic obstructive pulmonary disease (*P*=.005). The nonsevere COVID-19 and non-COVID viral infection classes differed significantly in 12 features, including smoking habit (*P*<.001), fever (*P*<.001), and sore throat (*P*=.002). However, the nonsevere COVID-19 and non-COVID viral infection classes did not differ significantly in terms of comorbidities. The severe COVID-19 and non-COVID viral infection classes differed significantly in 16 out of 22 features, making these 2 classes the most distinct in terms of symptoms. These results showed that the prevalence of clinical features differed substantially between the classes. The complete z-test results for each clinical feature in each pair of classes are provided in Multimedia Appendix 1.

In addition, based on the ML RF analysis, the top 3 differentiating clinical features were fever, coughing, and old age (ie, >50 years). For fever and coughing, we used the non-COVID viral infection class as the reference and constructed 2×2 contingency tables for the nonsevere COVID-19 and non-COVID viral infection classes, and the severe COVID-19 and non-COVID viral infection classes. The odds ratios and 95% confidence intervals for the forest plot are shown in Figure 2. Compared to patients in the non-COVID viral infection class, patients in both the nonsevere and severe COVID-19 classes were more likely to develop fever (ie, >37°C). In addition, based on the forest plot, patients with severe COVID-19 also experienced more fevers than patients with nonsevere COVID-19. Therefore, fever was one of the major determining factors of differentiating between multiple classes. Furthermore, patients with nonsevere COVID-19 (*P*<.001) and patients with severe COVID-19 (*P*<.001) reported significantly less coughing than the patients with non-COVID viral infection (Figure 2). There were no statistically significant differences between the nonsevere and severe COVID-19 classes in terms of clinical features. With regard to the old age feature, we included the severe COVID-19, nonsevere COVID-19, and noninfected healthy classes in the analysis because the prevalence of old age in the noninfected healthy class was not 0. The forest plot for this analysis is shown in Figure 2. Patients with severe COVID-19 were significantly older than patients with non-COVID viral infection, while patients with nonsevere COVID-19 and noninfected healthy individuals were younger than patients with non-COVID viral infection. These differences in clinical features between the 4 classes could pave the way toward a data-driven ML model.

Figure 2. Forest plot of the top 3 differentiating clinical features. Viral pneumonia was used as the reference class during comparisons and the calculation of odds ratios. The noninfected healthy class had no individuals with fevers or coughs. Therefore, these individuals were not included in the first 2 graphs (ie, the left and middle graphs). The error bars represent variation in estimated odds ratios, not the original feature variations.

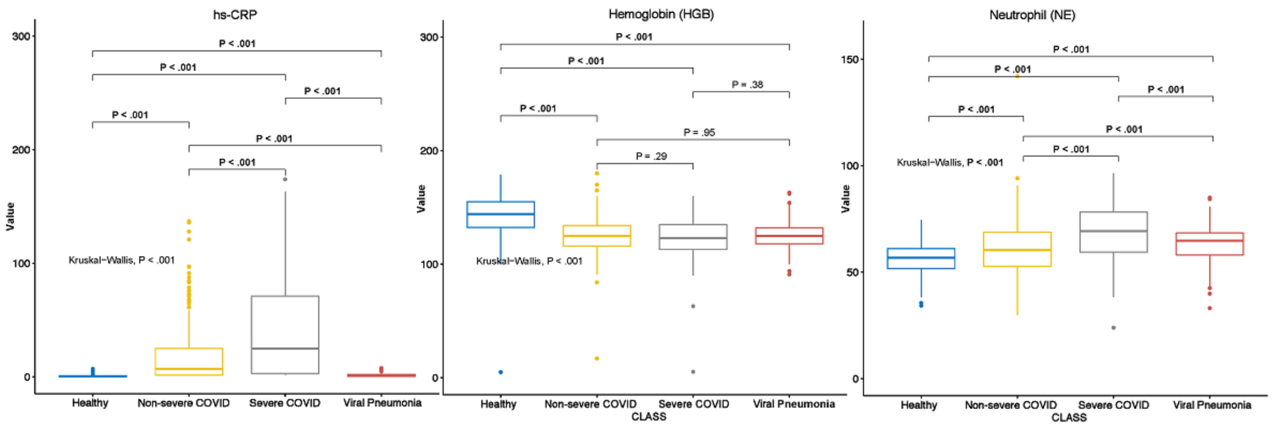


Differences in Lab Testing Features Between the 4 Classes

With regard to the continuous lab testing features, we calculated and compared the exact distributions among the 4 classes. The boxplots for each lab testing feature across the 4 classes are provided in Multimedia Appendix 4. In general, the 4 classes differed substantially across many lab testing features. Based on the 2-sided Kolmogorov-Smirnov test results, the nonsevere and severe COVID-19 classes were only similar in hemoglobin level (HGB *P*=.74) and platelet count (PLT *P*=.61). These 2 classes differed significantly in the remaining 8 lab testing features (WBC *P*=.02; NE% *P*<.001; NE *P*<.001; LY% *P*<.001; LY *P*=.002; CRP *P*<.001; TBIL *P*=.001; CREA *P*<.001;) . In other words, the lab testing features of patients with severe or nonsevere COVID-19 had distinct distributions. Similarly, the nonsevere COVID-19 and noninfected healthy classes were only similar in creatine level; the nonsevere COVID-19 and non-COVID viral infection classes were only similar in hemoglobin level (*P*=.65), platelet count (*P*=.14), and total bilirubin level (*P*=.09); the severe COVID-19 and noninfected healthy classes were only similar in total bilirubin level (*P*=.24); the severe COVID-19 and non-COVID viral infection classes were only similar in hemoglobin level (*P*=.11) and neutrophil count (*P*=.08); and the non-COVID viral infection and noninfected healthy classes were only similar in white blood cell count (*P*=.70). The complete Kolmogorov-Smirnov test results for each lab testing feature in each pair of classes are provided in Multimedia Appendix 1.

Based on the RF model, the 3 most influential differentiating features were C-reactive protein level, hemoglobin level, and neutrophil count. The distribution of C-reactive protein level among the 4 classes are provided in the boxplot in Figure 3. In addition to the Kolmogorov-Smirnov test, which did not account for multiple comparisons between classes, further pairwise comparisons were performed with the nonparametric Kruskal-Wallis H test. Each of the 6 pairs used in the Kruskal-Wallis H test, as well as the overall Kruskal-Wallis test, showed significant differences between each class. The distribution of hemoglobin levels is shown in Figure 3. Although the noninfected healthy class differed significantly from the nonsevere COVID-19, severe COVID-19, and non-COVID viral infection class in terms of hemoglobin level, the other 3 pairs did not show statistically significant differences in lab testing features. The distribution of neutrophil count is shown in Figure 3. All pairwise comparisons and the overall Kruskal-Wallis test showed significant differences between classes in terms of lab testing features.

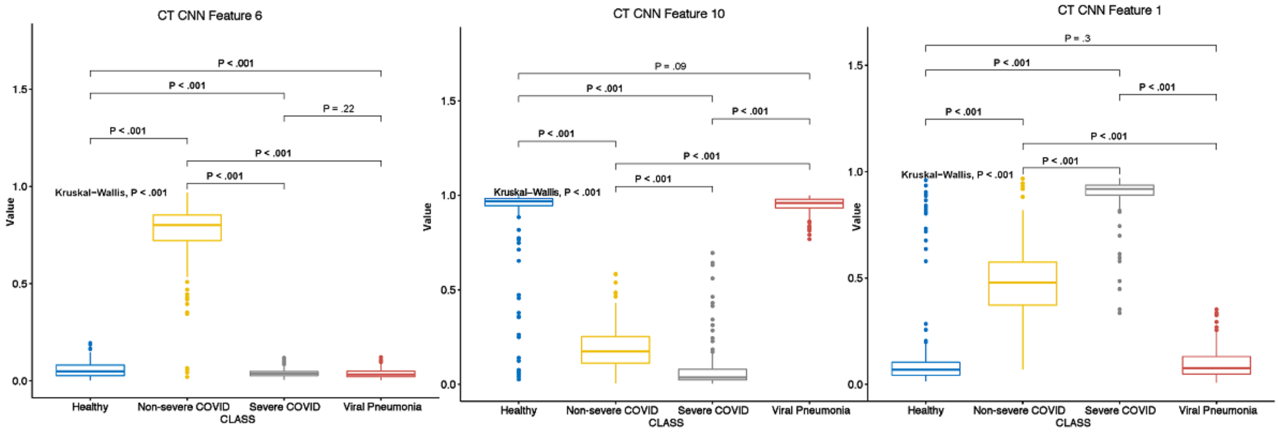
Figure 3. Multiple comparisons of the top differentiating lab testing features. hs-CRP: high-sensitivity C-reactive protein.



CT Differences Between the 4 Classes Based on High-Level CNN Features

We analyzed the FC2 layer features from the ResNet CNN in relation to the 4 classes. The corresponding boxplot is shown in Multimedia Appendix 5. The 2-sided Kolmogorov-Smirnov tests showed significant differences between every pair of classes in almost all 10 CT features in the FC2 layer. The only exceptions were feature 6 (ie, CNN6) between the severe COVID-19 and non-COVID viral infection classes and features 1, 4, and 5 between the noninfected healthy and non-COVID viral infection classes (Multimedia Appendix 6). Based on the RF model results, features 1, 6, and 10 were the 3 most critical features in the FC2 layer with regard to multinomial classification. Further Kruskal-Wallis tests were performed for these 3 features, and the results are shown in Figure 4. These results showed that developing an accurate classifier based on the CNN representation of high-level features is possible.

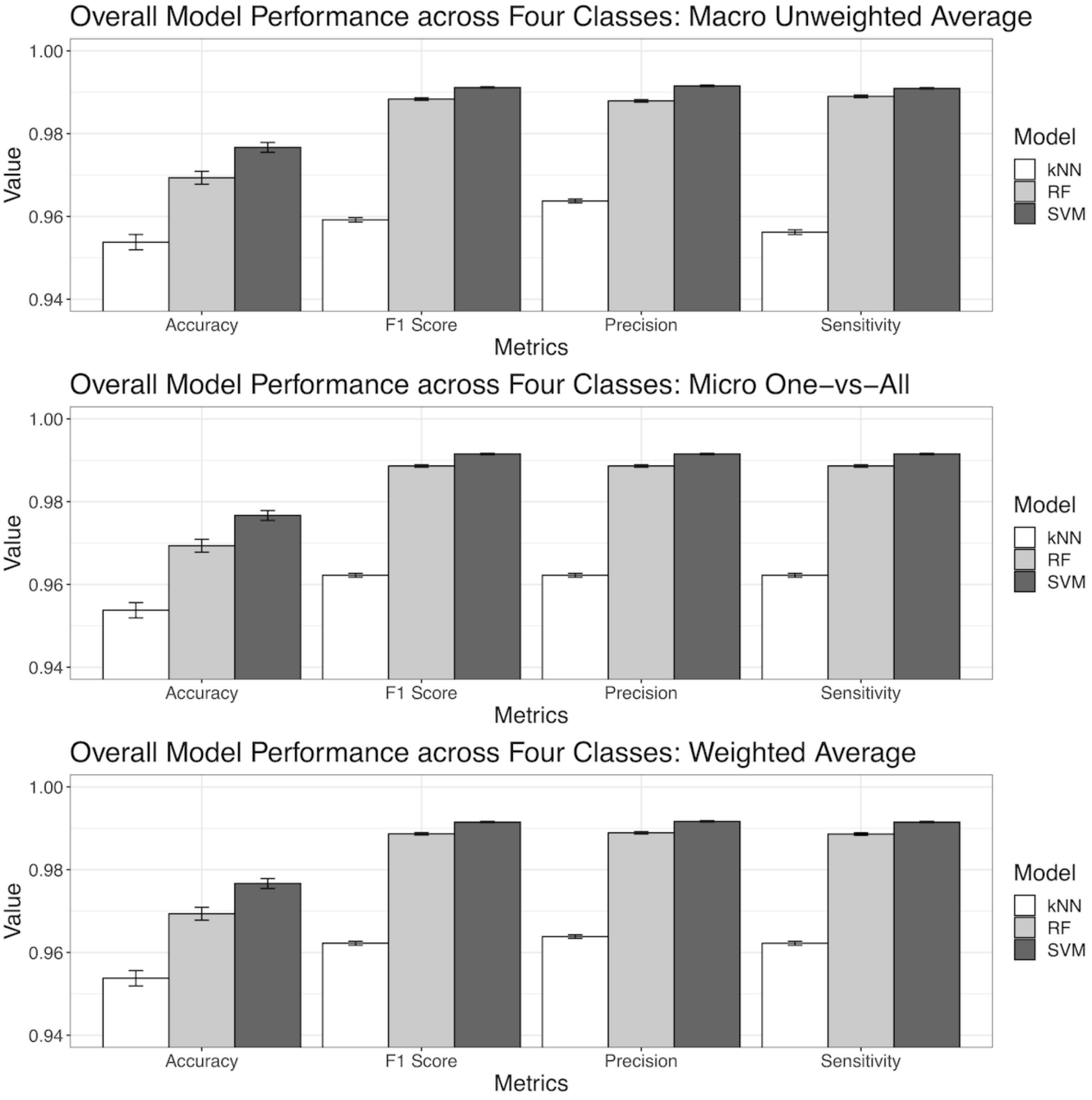
Figure 4. Multiple comparisons of the top differentiating CT features in the CNN. CNN: convolutional neural network; CT: computed tomography.



Accurate Multimodal Model for COVID-19 Multinomial Classification

We developed and validated 3 different types of ML models, as follows: the kNN, RF, and SVM models. With regard to training data, the average overall multimodal classification accuracy of the kNN, RF, and SVM models was 96.2% (SE 0.5%), 99.8% (SE 0.3%), and 99.2% (SE 0.2%), respectively. With regard to test data, the average overall multimodal classification accuracy of the 3 models was 95.4% (SE 0.2%), 96.9% (SE 0.2%), and 97.7% (SE 0.1%), respectively (Figure 5). These 3 models also achieved consistent and high performance across all 4 classes based on the different approaches for calculating the overall performance, including the micro approach (ie, the one-vs-all approach), macro approach (ie, unweighted averages across all 4 classes), and weighted average approach (ie, based on percentage of each class in the entire sample). It should be noted that overall accuracy did not depend on sample size, so there was only 1 approach for calculating accuracy. The F1 score, sensitivity, and precision were quantified via each approach (ie, the micro, macro, and weighted average approaches). The F1 scores that were calculated using the macro approach were 95.9% (SE 0.1%), 98.8% (SE<0.1%), and 99.1% (SE<0.1%) for the kNN, RF, and SVM models, respectively. The F1 scores that were calculated using the micro approach was 96.2% (SE<0.1%), 98.8% (SE<0.1%), and 99.2% (SE<0.1%) for the kNN, RF, and SVM models, respectively. The F1 scores calculated using the weighted average approach was 96.2% (SE<0.1%), 98.9% (SE<0.1%), and 99.2% (SE<0.1%) for the kNN, RF, and SVM models, respectively. The differences in F1 scores based on the different approaches (ie, the micro, macro, and weighted average approaches) were minimal (Figure 5). In addition, the differences in F1 scores across the different ML models (Figure 5) were also not significant. Similarly, model sensitivity and precision were all >95% for all ML model types and all approaches for calculating the performance metric. The complete overall performance metrics for the 3 different evaluation approaches and 3 ML models are presented in Multimedia Appendix 7.

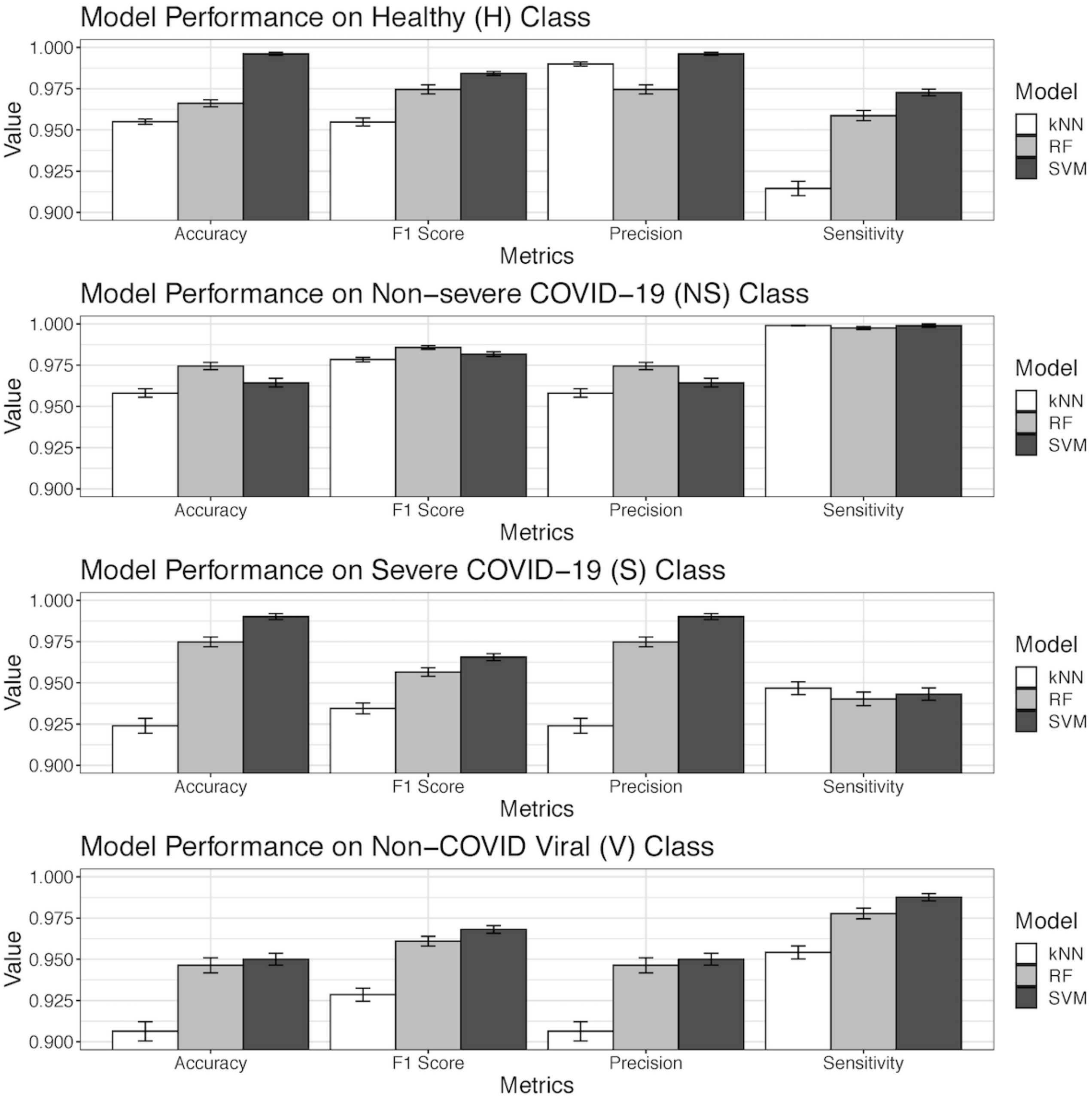
Figure 5. The overall performance of machine learning models across the 4 classes. Model performance was based on the prediction of unseen testing data (ie, the 20% of the original data), not on the 80% of the original data that were used to develop the model. kNN: k nearest neighbor; RF: random forest; SVM: support vector machine.



After examining the performance metrics across the 3 different types of ML models, it was clear that the SVM model consistently had the best performance with regard to all metrics, followed by the RF model, though the difference was almost indistinguishable. The kNN model had about a 1%-3% deficiency in performance compared to the other 2 models. It should be noted that the kNN model also had an accuracy, F1 score, sensitivity, and precision of at least 95%. Therefore, the kNN model was only bested by 2 even more competitive models. Furthermore, the relatively small standard errors demonstrated that the ML models were robust against different randomly sampled inputs (Multimedia Appendix 7).

With regard to each individual class, the noninfected healthy class had a 95.2%-99.9% prediction accuracy, 95.5%-98.4% F1 score, 91.4%-97.3% sensitivity, and 97.5%-99.9% precision in the testing set, depending on the specific ML model used. It should be noted these are ranges, not standard errors, as shown in Figure 6. The approach to computing class-specific model performance was the one-vs-all approach. With regard to the nonsevere COVID-19 class, ML models achieved a 95.8%-97.4% accuracy, 97.8%-98.6% F1 score, 99.8%-99.9% sensitivity, and 95.8%-97.4% precision. With regard to the severe COVID-19 class, ML models achieved a 92.4%-99.0% accuracy, 93.4%-96.6% F1 score, 94.3%-94.7% sensitivity, and 92.4%-99.0% precision. With regard to the non-COVID viral pneumonia infection class, ML models achieved a 90.6%-95.0% accuracy, 92.9%-96.8% F1 score, 95.4%-98.8% sensitivity, and 90.6%-95.0% precision. The non-COVID viral infection class was relatively more challenging to differentiate from the other 3 classes, but the difference was not substantial. Therefore, the potential clinical use of the ML models is still justified. Similar to the results of overall model performance (Figure 5), class-specific performance metrics also had relatively small standard errors, indicating that the training of models was consistent and robust against randomly selected inputs. Except for a few classes and model performance metrics, the SVM model performed slightly better than the RF and kNN models. The complete class-specific results are shown in Figure 6. The complete class-specific performance metrics across the 3 ML models are shown in Multimedia Appendix 8.

Figure 6. Class-specific performance of machine learning models. kNN: k nearest neighbor; RF: random forest; SVM: support vector machine.



All 3 ML multinomial classification models, which were based on different computational techniques, had consistently high overall performance (Figure 5, Table S3) and high performance for each specific class (Figure 6, Multimedia Appendix 8). Of

the 3 types of ML models developed and evaluated, the SVM model was marginally better than the RF and kNN models. As a result, the ML multinomial classification models were able to accurately differentiate between the 4 classes all at once, provide accurate and detailed class-specific predictions, and act as reliable decision-making tools for clinical diagnostic support and the triaging of patients with suspected COVID-19, who might or might not be infected with a clinically similar type of virus other than SARS-CoV-2.

In addition to the multimodal classification that incorporated all 3 different feature sets (ie, binary clinical, continuous lab testing, and CT features in the ResNet CNN; Figure 1), we also tested how each specific feature modality performed without feature fusion (ie, unimodality). By using each of the 23 symptom features alone, the RF, kNN, and SVM models achieved an average accuracy of 74.5% (SE 0.3%), 73.3% (SE 0.3%), and 75.5% (SE 0.3%) with the testing set, respectively. By using each of the 10 lab testing features alone, the RF, kNN, and SVM models achieved an average accuracy of 67.7% (SE 0.4%), 56.2% (SE 0.4%), and 59.5% (SE 0.3%) with the testing set, respectively.

The overall accuracy of the CNN with CT scan data alone was 90.8% (SE 0.3%) across the 4 classes. With regard to each pair of classes, the CNN was able to accurately differentiate between the severe COVID-19 and noninfected healthy classes with 99.9% (SE<0.1%) accuracy, the non-COVID viral infection and noninfected healthy classes with 99.2% (SE 0.1%) accuracy, the severe COVID-19 and nonsevere COVID-19 classes with 95.4% (SE 0.1%) accuracy, and the non-severe COVID-19 and noninfected healthy classes with 90.3% (SE 0.2%) accuracy. However, by using CT features alone (ie, without feature late fusion), the CNN could only differentiate between the non-COVID viral infection and nonsevere COVID-19 classes with 84.9% (SE 0.2%) accuracy, and the non-COVID viral infection and severe COVID-19 with 74.2% (SE 0.2%) accuracy in the testing set.

Substantial performance boosts were gained by combining input features from the different feature modalities and performing multimodal classification, instead of using a single-feature modality alone. A 15%-42% increase in prediction accuracy with the testing set was achieved compared to the single-modality models. It should be noted that the RF, SVM, and kNN models were technically distinct ML models. However, the performance differences between these 3 distinct ML models were marginal, based on the multimodal features. Therefore, we concluded that the high performance in COVID-19 classification in this study (Figures 5 and 6) was largely due to multimodal feature late fusion, not due to the specific type of ML model.

Gini impurity scores derived from the RF model identified major contributing factors that differentiated the 4 classes. With regard to clinical feature modality, the top 3 most influential features were fever, coughing, and old age (ie, >50 years). The forest plots of odds ratios for these features are provided in Figure 2, which shows the exact influence that these features had across classes. With regard to lab testing features, the top 3 most influential features, in descending order, were high-sensitivity C-reactive protein level, hemoglobin level, and absolute neutrophil count. The distribution of these 3 features across the 4 classes and the results of multiple comparisons are shown in Figure 3. Although high-sensitivity C-reactive protein level is a known factor for COVID-19 severity and prognosis [42], we showed that it could also differentiate patients with COVID-19 from patients with non-COVID viral pneumonia and healthy individuals. In addition, we learned that different hemoglobin and neutrophil levels were novel features for accurately distinguishing between patients with clinical COVID-19, patients with non-COVID viral pneumonia, and healthy individuals. These results shed light on which set of clinical and lab testing features are the most critical in identifying COVID-19, which will help guide clinical practice. With regard to the CT features extracted from the CNN, the RF models identified the top 3 influential features, which were CT features 6, 10, and 1 in the 10-element FC2 layer (Figure 4). Although the actual clinical interpretation of CT features was not clear at the time of this study due to the nature of DL models, including the ResNet CNN applied in this study, these features showed promise in accurately differentiating between multinomial classes all at once via CT scans, instead of training several CNNs for binary classifications between each class pair. Future research might reveal the clinical relevance of these features in a more interpretable way with COVID-19 pathology data.

Discussion

Principal Findings

In this study, we provided a more holistic perspective to characterizing COVID-19 and accurately differentiating COVID-19, especially nonsevere COVID-19, from other clinically similar viral pneumonias and noninfections. The human body is an integrated and systemic entity. When the body is infected by pathogens, clinical consequences can be detected not only with biomedical imaging features (eg, CT scan features), but also with other features, such as lab testing results for blood biochemistry [20,43]. A single-feature modality might not reveal the full clinical consequences and provide the best predictive power for COVID-19 detection and classification, but the synergy of multiple modalities exceeds the power of any single modality. Currently, multimodality medical data can be effectively stored, transferred, and exchanged with electronic health record systems. The economic cost of acquiring clinical and lab testing modality data are lower than the economic cost of acquiring current confirmatory qRT-PCR data. Availability and readiness are also advantages that these modalities have over qRT-PCR, which currently has a long turnaround time. This study harnessed the power of multimodality medical information for an emerging pandemic, for which confirmatory molecular tests have reliability and availability issues across time and space. This study's novel analytical framework can be used to prepare for incoming waves of disease epidemics in the future, when clinicians' experience and understanding with the disease may vary substantially.

Upon the further examination of comprehensive patient symptom data, we believed that our current understanding and definition of asymptomatic COVID-19 would be inadequate.

Of the 214 patients with nonsevere COVID-19, 60 (28%) had no fever (ie, <37°C), 78 (36.4%) did not experience coughing, 141 (65.9%) did not feel chest congestion and pain, and 172 (80.4%) did not report having a sore throat upon admission. Additionally, there were 10 (4.7%) patients with confirmed COVID-19 in the nonsevere COVID-19 class who did not present with any of these common symptoms and could be considered patients with asymptomatic COVID-19. Even after considering headache, muscle pain, and fatigue, there were still 4 (1.9%) patients who did not show symptoms related to typical respiratory diseases. Of these 4 patients, 1 (25%) had diarrhea upon admission. Therefore, using symptom features alone is not sufficient for detecting and differentiating patients with asymptomatic COVID-19. Nevertheless, all asymptomatic patients were successfully detected via our model, and no false negatives were observed. This finding shows the incompleteness of the current definition and understanding of asymptomatic COVID-19, and the potential power that nontraditional analytical tools have for identifying these patients.

Based on this perspective, we developed a comprehensive end-to-end analytical framework that integrated both high-dimensional biomedical imaging data and low-dimensional clinical and lab testing data. CT scans were first processed with DL CNNs. We developed a customized ResNet CNN architecture with 2 FC layers before the final output layer. We then used the second FC layer as the low-dimensional representation of the original high-dimensional CT data. In other words, a CNN was applied first for dimensional reduction. The feature fusion of CT (ie, represented by the FC layers), clinical, and lab testing feature modalities demonstrated feasibility and high accuracy in differentiating between the nonsevere COVID-19, severe COVID-19, non-COVID viral pneumonia, and noninfected healthy classes all at once. The consistent high performance across the 3 different types of ML models (ie, the RF, SVM, and kNN models), as well as the substantial performance boost from using a single modality, further unleashed the hidden power of feature fusion for different biomedical feature modalities. Compared to the accuracy of using any single-feature modality alone (60%-80%), the feature fusion of multimodal biomedical data substantially boosted prediction accuracy (>97%) in the testing set.

We compared the performance of our model, which was based on the multimodal biomedical data of 683 participants, against the performance of state-of-the-art benchmarks in COVID-19 classification studies. A DL study that involved thoracic CT scans for 87 participants claimed to have >99% accuracy [37], and another study with 200 participants claimed to have 86%-99% accuracy in differentiating between individuals with and without COVID-19 [36]. Another study reported a 95% area under the curve for differentiating between COVID-19 and other community-acquired pneumonia diseases in 3322 participants [39]. Furthermore, a 92% area under the curve was achieved in a study of 905 participants with and without COVID-19 by using multimodal CT, clinical, and lab testing information [44]. A study that used CT scans to differentiate between 3 multinomial classes (ie, the COVID [no clinical state information], non-COVID viral pneumonia, and healthy classes) achieved an 89%-96% accuracy based on a total of 230 participants [38]. In addition, professionally trained human radiologists have achieved a 60%-83% accuracy in differentiating COVID-19 from other types of community-acquired pneumonia diseases [45]. Therefore, the performance of our model is on par with, or superior to, the performance of these benchmark models and exceeds the performance of human radiologists. Moreover, previous studies have generally focused on differentiating patients with COVID-19 from individuals without COVID-19 or patients with other types of pneumonia. In other words, the current COVID-19 classification models are mostly binary classifiers. Our study not only detected COVID-19 in healthy individuals, but also addressed the more important clinical issue of differentiating COVID-19 from other viral infections. Our study also distinguished between different COVID-19 clinical states (ie, severe vs nonsevere). Therefore, our study provides a novel and effective breakthrough for clinical applications, not just incremental improvements for existing ML models.

The success of this study sheds light on many other disease systems that use multimodal biomedical data inputs. Specifically, the feature fusion of high- and low-dimensional biomedical data modalities can be applied to more feature modalities, such as individual-level high-dimensional "-omics" data. Currently, a study on the genome-wide association between individual single nucleotide polymorphisms and COVID-19 susceptibility has revealed several target loci that are involved in COVID-19 pathology. Following a similar approach, we may also conduct another study, in which we first carry out the dimensional reduction of "-omics" data, and then perform data fusion with other low-dimensional modalities [46-48].

With regard to classification, this study adopts a hybrid of DL (ie, CNN) and ML (ie, RF, SVM, and kNN ML) models via feature late fusion. By using various data-driven methods, we avoided the potential cause-effect pitfall and focused directly on the more important clinical question. For instance, many comorbidities, such as diabetes [49,50] and cardiovascular diseases [51,52], are strongly associated with the occurrence of severe COVID-19. It is still unclear whether diabetes or reduced kidney function causes severe COVID-19, whether SARS-CoV-2 infection worsens existing diabetes, or whether diabetes and COVID-19 actually mutually influence each other and result in undesirable clinical prognoses. Future studies can use data-driven methods to further investigate the causality of comorbidities and COVID-19.

There are some limitations in this study and potential improvements for future research. For instance, to perform multinomial classification across the 4 classes, we had to discard a lot of features, especially those in the lab testing modality. The non-COVID viral pneumonia class used a different electronic health record system that collected different lab testing features from participants in Wuhan (ie, participants in the severe COVID-19, nonsevere COVID-19, and noninfected healthy classes). Many lab testing features were able to accurately differentiate between severe and nonsevere COVID-19 in our preliminary study, such as high-sensitivity Troponin I level, D-dimer level, and lactate dehydrogenase level. However, these features were not present, or largely missing, in the non-COVID viral infection class. Eventually, only 10 lab testing features were included, which is small compared to the average of 20-30 features that are usually available in different electronic health record systems. This is probably the reason why the lab testing feature modality alone was not able to provide accurate classifications (ie, the highest accuracy achieved was 67.7% with the RF model) across all 4 classes in this study. In addition, although we had a reasonably large participant pool of 638 individuals, more participants are needed to further validate the findings of this study.

Another potential practical pitfall was that not all feature modalities were readily available at the same time for feature fusion and multimodal classification. With regard to single-modality features, CT had the best performance in generating accurate predictions. However, CT is usually performed in the radiology department. Lab testing may be outsourced, and obtaining lab test results takes time. Consequently, there might be lags in data availability among different feature modalities. We believe that when multimodal features are not available all at once, single-modality features can be used to perform first-round triaging. Multimodal features are needed when accuracy is a must.

It should be noted that although the participants in this study came from different health care facilities, the majority of them were of Chinese Han ethnicity. The biomedical features among the different COVID-19 and non-COVID classes may be different in people of other races and ethnicities, or people with other confounding factors. The cross-validation of the findings in this study based on other ethnicity groups and larger sample sizes is needed for future research.

This study used a common CNN architecture (ie, a ResNet). The 10 CT features extracted from the FC2 layer of the ResNet were used to match the dimensionality of the other 2 low-dimensional feature modalities. Future research on different disease systems can explore and compare other architectures that use different biomedical imaging data (eg, CT, X-ray, and histology data). The actual dimensionality of the FC2 layer can also be optimized to deliver better performance. Finally, this study presented the results of individual classification models. To achieve even higher performance, the combination of multiple models can be explored in future studies.

Conclusion

In summary, different biomedical information across different modalities, such as clinical information, lab testing results, and CT scans, work synergistically to reveal the multifaceted nature of COVID-19 pathology. Our ML and DL models provided a feasible technical method for working directly with multimodal biomedical data and differentiating between patients with severe COVID-19, patients with nonsevere COVID-19, patients with non-COVID viral infection, and noninfected healthy individuals at the same time, with >97% accuracy.

Acknowledgments

This study is dedicated to the frontline clinicians and other supporting personnel who have fought against COVID-19 worldwide. This study is supported by the National Science Foundation for Young Scientists of China (81703201, 81602431, and 81871544), the North Carolina Biotechnology Center Flash Grant on COVID-19 Clinical Research (2020-FLG-3898), the Natural Science Foundation for Young Scientists of Jiangsu Province (BK20171076, BK20181488, BK20181493, and BK20201485), the Jiangsu Provincial Medical Innovation Team (CXTDA2017029), the Jiangsu Provincial Medical Youth Talent program (QNRC2016548 and QNRC2016536), the Jiangsu Preventive Medicine Association program (Y2018086 and Y2018075), the Lifting Program of Jiangsu Provincial Scientific and Technological Association, and the Jiangsu Government Scholarship for Overseas Studies.

Authors' Contributions

JL (liu_jie0823@163.com), BZ (zhubl@jscdc.cn), and SC (schen56@uncc.edu) serve as corresponding authors of this study equally.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic, clinical, and lab testing results of the 4 Classes.
[DOCX File , 18 KB - jmir_v23i1e25535_app1.docx]

Multimedia Appendix 2
Architecture of the customized ResNet-18 CNN and sample computed tomography scans of the 4 classes. CNN: convolutional neural networkl ResNet: residual neural network.
[PNG File , 433 KB - jmir_v23i1e25535_app2.png]

Multimedia Appendix 3
Comparison of clinical features across the 4 classes.
[PNG File , 122 KB - jmir_v23i1e25535_app3.png]

Multimedia Appendix 4
Comparison of lab testing features across the 4 classes.
[PNG File , 97 KB - jmir_v23i1e25535_app4.png]

Multimedia Appendix 5
Computed tomography features extracted via a deep learning convolutional neural network and compared across the 4 classes.
[PNG File , 131 KB - jmir_v23i1e25535_app5.png]

Multimedia Appendix 6
z-test and Kolmogorov-Smirnov test results of significance for each biomedical feature among the 4 classes.
[DOCX File , 21 KB - jmir_v23i1e25535_app6.docx]

Multimedia Appendix 7
Overall machine learning model performance comparison.
[DOCX File , 15 KB - jmir_v23i1e25535_app7.docx]

Multimedia Appendix 8
Class-specific machine learning model performance comparison.
[DOCX File , 15 KB - jmir_v23i1e25535_app8.docx]

References

1. Coronavirus Disease 2019 (COVID-19) Situation Report - 203. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200810-covid-19-sitrep-203.pdf?sfvrsn=aa050308_2 [accessed 2020-12-22]
2. Novel coronavirus pneumonia diagnosis and treatment plan. National Health Commission of China. 2020. URL: http://www.nhc.gov.cn/yzygj/s7652m/202003/a31191442e29474b98bfed5579d5af95.shtml [accessed 2020-12-22]
3. Xiao SY, Wu Y, Liu H. Evolving status of the 2019 novel coronavirus infection: Proposal of conventional serologic assays for disease diagnosis and infection monitoring. J Med Virol 2020 May;92(5):464-467 [FREE Full text] [doi: 10.1002/jmv.25702] [Medline: 32031264]
4. Wang Y, Kang H, Liu X, Tong Z. Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. J Med Virol 2020 Jun;92(6):538-539 [FREE Full text] [doi: 10.1002/jmv.25721] [Medline: 32096564]
5. Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. Am J Infect Control. Epub ahead of print 2020 Jul 10 [FREE Full text] [doi: 10.1016/j.ajic.2020.07.011] [Medline: 32659413]
6. Zhou L, Li Z, Zhou J, Li H, Chen Y, Huang Y, et al. A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis. IEEE Trans Med Imaging 2020 Aug;39(8):2638-2652. [doi: 10.1109/TMI.2020.3001810] [Medline: 32730214]
7. Fu Y, Zhu R, Bai T, Han P, He Q, Jing M, et al. Clinical Features of COVID-19-Infected Patients With Elevated Liver Biochemistries: A Multicenter, Retrospective Study. Hepatology. Epub ahead of print 2020 Jun 30 [FREE Full text] [doi: 10.1002/hep.31446] [Medline: 32602604]
8. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 2020 Apr;20(4):425-434 [FREE Full text] [doi: 10.1016/S1473-3099(20)30086-4] [Medline: 32105637]
9. Ojha V, Mani A, Pandey NN, Sharma S, Kumar S. CT in coronavirus disease 2019 (COVID-19): a systematic review of chest CT findings in 4410 adult patients. Eur Radiol 2020 Nov;30(11):6129-6138 [FREE Full text] [doi: 10.1007/s00330-020-06975-7] [Medline: 32474632]
10. Lyu P, Liu X, Zhang R, Shi L, Gao J. The Performance of Chest CT in Evaluating the Clinical Severity of COVID-19 Pneumonia: Identifying Critical Cases Based on CT Characteristics. Invest Radiol 2020 Jul;55(7):412-421 [FREE Full text] [doi: 10.1097/RLI.0000000000000689] [Medline: 32304402]
11. Brooks M. How accurate is self-testing? New Sci 2020 May 16;246(3282):10 [FREE Full text] [doi: 10.1016/S0262-4079(20)30909-X] [Medline: 32501335]
12. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
13. Truog RD, Mitchell C, Daley GQ. The Toughest Triage - Allocating Ventilators in a Pandemic. N Engl J Med 2020 May 21;382(21):1973-1975. [doi: 10.1056/NEJMp2005689] [Medline: 32202721]

14. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology 2020 Aug;296(2):E115-E117 [FREE Full text] [doi: 10.1148/radiol.2020200432] [Medline: 32073353]
15. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
16. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study. JMIR Mhealth Uhealth 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: 10.2196/19822] [Medline: 32516750]
17. Nair A, Rodrigues JCL, Hare S, Edey A, Devaraj A, Jacob J, et al. A British Society of Thoracic Imaging statement: considerations in designing local imaging diagnostic algorithms for the COVID-19 pandemic. Clin Radiol 2020 May;75(5):329-334 [FREE Full text] [doi: 10.1016/j.crad.2020.03.008] [Medline: 32265036]
18. Sun Y, Koh V, Marimuthu K, Ng O, Young B, Vasoo S, National Centre for Infectious Diseases COVID-19 Outbreak Research Team. Epidemiological and Clinical Predictors of COVID-19. Clin Infect Dis 2020 Jul 28;71(15):786-792 [FREE Full text] [doi: 10.1093/cid/ciaa322] [Medline: 32211755]
19. Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. J Med Syst 2020 Jul 01;44(8):135 [FREE Full text] [doi: 10.1007/s10916-020-01597-4] [Medline: 32607737]
20. Daniells JK, MacCallum HL, Durrheim DN. Asymptomatic COVID-19 or are we missing something? Commun Dis Intell (2018) 2020 Jul 09;44:1-5 [FREE Full text] [doi: 10.33321/cdi.2020.44.55] [Medline: 32640950]
21. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. N Engl J Med 2020 May 28;382(22):2158-2160 [FREE Full text] [doi: 10.1056/NEJMe2009758] [Medline: 32329972]
22. Shi F, Yu Q, Huang W, Tan C. 2019 Novel Coronavirus (COVID-19) Pneumonia with Hemoptysis as the Initial Symptom: CT and Clinical Features. Korean J Radiol 2020 May;21(5):537-540 [FREE Full text] [doi: 10.3348/kjr.2020.0181] [Medline: 32174057]
23. Li X, Fang X, Bian Y, Lu J. Comparison of chest CT findings between COVID-19 pneumonia and other types of viral pneumonia: a two-center retrospective study. Eur Radiol 2020 Oct;30(10):5470-5478 [FREE Full text] [doi: 10.1007/s00330-020-06925-3] [Medline: 32394279]
24. Altmayer S, Zanon M, Pacini GS, Watte G, Barros MC, Mohammed TL, et al. Comparison of the computed tomography findings in COVID-19 and other viral pneumonia in immunocompetent adults: a systematic review and meta-analysis. Eur Radiol 2020 Dec;30(12):6485-6496 [FREE Full text] [doi: 10.1007/s00330-020-07018-x] [Medline: 32594211]
25. Qu J, Chang LK, Tang X, Du Y, Yang X, Liu X, et al. Clinical characteristics of COVID-19 and its comparison with influenza pneumonia. Acta Clin Belg 2020 Oct;75(5):348-356. [doi: 10.1080/17843286.2020.1798668] [Medline: 32723027]
26. Ooi EE, Low JG. Asymptomatic SARS-CoV-2 infection. Lancet Infect Dis 2020 Sep;20(9):996-998 [FREE Full text] [doi: 10.1016/S1473-3099(20)30460-6] [Medline: 32539989]
27. Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. Sci Rep 2019 Apr 23;9(1):6381 [FREE Full text] [doi: 10.1038/s41598-019-42294-8] [Medline: 31011155]
28. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
29. Song S, Wu F, Liu Y, Jiang H, Xiong F, Guo X, et al. Correlation Between Chest CT Findings and Clinical Features of 211 COVID-19 Suspected Patients in Wuhan, China. Open Forum Infect Dis 2020 Jun;7(6):ofaa171 [FREE Full text] [doi: 10.1093/ofid/ofaa171] [Medline: 32518804]
30. Wu J, Wu X, Zeng W, Guo D, Fang Z, Chen L, et al. Chest CT Findings in Patients With Coronavirus Disease 2019 and Its Relationship With Clinical Features. Invest Radiol 2020 May;55(5):257-261 [FREE Full text] [doi: 10.1097/RLI.0000000000000670] [Medline: 32091414]
31. Meng H, Xiong R, He R, Lin W, Hao B, Zhang L, et al. CT imaging and clinical course of asymptomatic cases with COVID-19 pneumonia at admission in Wuhan, China. J Infect 2020 Jul;81(1):e33-e39 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.004] [Medline: 32294504]
32. Cheng Z, Lu Y, Cao Q, Qin L, Pan Z, Yan F, et al. Clinical Features and Chest CT Manifestations of Coronavirus Disease 2019 (COVID-19) in a Single-Center Study in Shanghai, China. AJR Am J Roentgenol 2020 Jul;215(1):121-126. [doi: 10.2214/AJR.20.22959] [Medline: 32174128]
33. Zhu Y, Liu YL, Li ZP, Kuang JY, Li XM, Yang YY, et al. Clinical and CT imaging features of 2019 novel coronavirus disease (COVID-19). J Infect. Epub ahead of print 2020 Mar 03 [FREE Full text] [doi: 10.1016/j.jinf.2020.02.022] [Medline: 32142928]
34. Baltrusaitis T, Ahuja C, Morency L. Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans Pattern Anal Mach Intell 2019 Feb;41(2):423-443. [doi: 10.1109/TPAMI.2018.2798607] [Medline: 29994351]
35. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, et al. Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious

Diseases Society of America. Am J Respir Crit Care Med 2019 Oct 01;200(7):e45-e67 [FREE Full text] [doi: 10.1164/rccm.201908-1581ST] [Medline: 31573350]

36. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput Biol Med 2020 Jun;121:103795 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103795] [Medline: 32568676]
37. Ko H, Chung H, Kang WS, Kim KW, Shin Y, Kang SJ, et al. COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep Learning Framework With a Single Chest CT Image: Model Development and Validation. J Med Internet Res 2020 Jun 29;22(6):e19569 [FREE Full text] [doi: 10.2196/19569] [Medline: 32568730]
38. Hu S, Gao Y, Niu Z, Jiang Y, Li L, Xiao X, et al. Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification From CT Images. IEEE Access 2020;8:118869-118883. [doi: 10.1109/ACCESS.2020.3005510]
39. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020 Aug;296(2):E65-E71 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
40. Chen Y, Ouyang L, Bao FS, Li Q, Han L, Zhu B, et al. An Interpretable Machine Learning Framework for Accurate Severe vs Non-severe COVID-19 Clinical Type Classification. medRxiv. Preprint posted online on May 22.. [doi: 10.1101/2020.05.18.20105841]
41. Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for image-based diagnosis of COVID-19. PLoS One 2020;15(6):e0235187 [FREE Full text] [doi: 10.1371/journal.pone.0235187] [Medline: 32589673]
42. Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and Laboratory Predictors of In-hospital Mortality in Patients With Coronavirus Disease-2019: A Cohort Study in Wuhan, China. Clin Infect Dis 2020 Nov 19;71(16):2079-2088 [FREE Full text] [doi: 10.1093/cid/ciaa538] [Medline: 32361723]
43. Ferrari D, Motta A, Strollo M, Banfi G, Locatelli M. Routine blood tests as a potential diagnostic tool for COVID-19. Clin Chem Lab Med 2020 Jun 25;58(7):1095-1099. [doi: 10.1515/cclm-2020-0398] [Medline: 32301746]
44. Mei X, Lee HC, Diao KY, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020 Aug;26(8):1224-1228 [FREE Full text] [doi: 10.1038/s41591-020-0931-3] [Medline: 32427924]
45. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, et al. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. Radiology 2020 Aug;296(2):E46-E54 [FREE Full text] [doi: 10.1148/radiol.2020200823] [Medline: 32155105]
46. Tárnok A. Machine Learning, COVID-19 (2019-nCoV), and multi-OMICS. Cytometry A 2020 Mar;97(3):215-216 [FREE Full text] [doi: 10.1002/cyto.a.23990] [Medline: 32142596]
47. Ray S, Srivastava S. COVID-19 Pandemic: Hopes from Proteomics and Multiomics Research. OMICS 2020 Aug;24(8):457-459. [doi: 10.1089/omi.2020.0073] [Medline: 32427517]
48. Arga KY. COVID-19 and the Futures of Machine Learning. OMICS 2020 Sep;24(9):512-514. [doi: 10.1089/omi.2020.0093] [Medline: 32511048]
49. Wicaksana AL, Hertanti NS, Ferdiana A, Pramono RB. Diabetes management and specific considerations for patients with diabetes during coronavirus diseases pandemic: A scoping review. Diabetes Metab Syndr 2020;14(5):1109-1120 [FREE Full text] [doi: 10.1016/j.dsx.2020.06.070] [Medline: 32659694]
50. Abdi A, Jalilian M, Sarbarzeh PA, Vlaisavljevic Z. Diabetes and COVID-19: A systematic review on the current evidences. Diabetes Res Clin Pract 2020 Aug;166:108347 [FREE Full text] [doi: 10.1016/j.diabres.2020.108347] [Medline: 32711003]
51. Madjid M, Safavi-Naeini P, Solomon SD, Vardeny O. Potential Effects of Coronaviruses on the Cardiovascular System: A Review. JAMA Cardiol 2020 Jul 01;5(7):831-840. [doi: 10.1001/jamacardio.2020.1286] [Medline: 32219363]
52. Matsushita K, Marchandot B, Jesel L, Ohlmann P, Morel O. Impact of COVID-19 on the Cardiovascular System: A Review. J Clin Med 2020 May 09;9(5):1407 [FREE Full text] [doi: 10.3390/jcm9051407] [Medline: 32397558]

Abbreviations

CT: computed tomography
CNN: convolutional neural network
DL: deep learning
DICOM: Digital Imaging and Communications in Medicine
FC: fully connected
GGO: ground-glass opacity
kNN: k-nearest neighbor
MERS: Middle East respiratory syndrome
ML: machine learning
qRT-PCR: quantitative real-time polymerase chain reaction
RF: random forest
ResNet: residual neural network
SARS: severe acute respiratory syndrome

SVM: support vector machine

Edited by G Eysenbach; submitted 05.11.20; peer-reviewed by K Ahmad, H Zhang; comments to author 26.11.20; revised version received 07.12.20; accepted 17.12.20; published 06.01.21.

Please cite as:
Xu M, Ouyang L, Han L, Sun K, Yu T, Li Q, Tian H, Safarnejad L, Zhang H, Gao Y, Bao FS, Chen Y, Robinson P, Ge Y, Zhu B, Liu J, Chen S
Accurately Differentiating Between Patients With COVID-19, Patients With Other Viral Infections, and Healthy Individuals: Multimodal Late Fusion Learning Approach
J Med Internet Res 2021;23(1):e25535
URL: http://www.jmir.org/2021/1/e25535/
doi:10.2196/25535
PMID:33404516

Original Paper

Investigating and Improving the Accuracy of US Citizens' Beliefs About the COVID-19 Pandemic: Longitudinal Survey Study

Aart van Stekelenburg[1], MSc; Gabi Schaap[1], PhD; Harm Veling[1], PhD; Moniek Buijzen[1,2], PhD

[1]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

[2]Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Aart van Stekelenburg, MSc
Behavioural Science Institute
Radboud University
Montessorilaan 3
Nijmegen, 6525 HR
Netherlands
Phone: 31 024 3615723
Email: a.vanstekelenburg@bsi.ru.nl

Abstract

Background: The COVID-19 infodemic, a surge of information and misinformation, has sparked worry about the public's perception of the coronavirus pandemic. Excessive information and misinformation can lead to belief in false information as well as reduce the accurate interpretation of true information. Such incorrect beliefs about the COVID-19 pandemic might lead to behavior that puts people at risk of both contracting and spreading the virus.

Objective: The objective of this study was two-fold. First, we attempted to gain insight into public beliefs about the novel coronavirus and COVID-19 in one of the worst hit countries: the United States. Second, we aimed to test whether a short intervention could improve people's belief accuracy by empowering them to consider scientific consensus when evaluating claims related to the pandemic.

Methods: We conducted a 4-week longitudinal study among US citizens, starting on April 27, 2020, just after daily COVID-19 deaths in the United States had peaked. Each week, we measured participants' belief accuracy related to the coronavirus and COVID-19 by asking them to indicate to what extent they believed a number of true and false statements (split 50/50). Furthermore, each new survey wave included both the original statements and four new statements: two false and two true statements. Half of the participants were exposed to an intervention aimed at increasing belief accuracy. The intervention consisted of a short infographic that set out three steps to verify information by searching for and verifying a scientific consensus.

Results: A total of 1202 US citizens, balanced regarding age, gender, and ethnicity to approximate the US general public, completed the baseline (T0) wave survey. Retention rate for the follow-up waves— first follow-up wave (T1), second follow-up wave (T2), and final wave (T3)—was high (≥85%). Mean scores of belief accuracy were high for all waves, with scores reflecting low belief in false statements and high belief in true statements; the belief accuracy scale ranged from –1, indicating completely inaccurate beliefs, to 1, indicating completely accurate beliefs (T0 mean 0.75, T1 mean 0.78, T2 mean 0.77, and T3 mean 0.75). Accurate beliefs were correlated with self-reported behavior aimed at preventing the coronavirus from spreading (eg, social distancing) (r at all waves was between 0.26 and 0.29 and all P values were less than .001) and were associated with trust in scientists (ie, higher trust was associated with more accurate beliefs), political orientation (ie, liberal, Democratic participants held more accurate beliefs than conservative, Republican participants), and the primary news source (ie, participants reporting CNN or Fox News as the main news source held less accurate beliefs than others). The intervention did not significantly improve belief accuracy.

Conclusions: The supposed infodemic was not reflected in US citizens' beliefs about the COVID-19 pandemic. Most people were quite able to figure out the facts in these relatively early days of the crisis, calling into question the prevalence of misinformation and the public's susceptibility to misinformation.

(J Med Internet Res 2021;23(1):e24069) doi:10.2196/24069

KEYWORDS

infodemic; infodemiology; misinformation; COVID-19 pandemic; belief accuracy; boosting; trust in scientists; political orientation; media use

Introduction

Public health crises tend to go hand in hand with information crises. The COVID-19 pandemic, which is taking many lives and is hospitalizing hundreds of thousands of people globally, is no exception. In the wake of the COVID-19 pandemic, we are seeing signs of a misinformation pandemic. Around the first peak of the coronavirus outbreak in the United States, the country with the highest COVID-19 death toll [1], about two-thirds of Americans said they had been exposed to at least some made-up news and information related to the virus [2]. Misinformation about the pandemic seems to have proliferated quickly, especially on social media [3]. The World Health Organization (WHO) has labelled this surge of information and misinformation about the COVID-19 pandemic an *infodemic* [4].

Countries and social media platforms are trying to tackle this infodemic in a number of ways. Several social media platforms, including Facebook and Twitter, have implemented new procedures to remove or label false and misleading content [5,6]. However, with the vast number of posts made to these platforms every day and the platforms' fear of infringing on free speech, the success of these procedures is limited (eg, [7]). A second strategy consists of surfacing trusted content, for instance, by referring people with questions to the WHO or to national health agencies, such as the Centers for Disease Control and Prevention in the United States and the National Epidemiology Center in Brazil. This approach might be hindered by government officials, including US president Donald Trump and Brazilian president Jair Bolsonaro, actually contributing to the spread of misinformation (eg, [8,9]). Considering this apparent infodemic, are people able to distinguish facts from fiction? And what correlates might enable or disable them in forming accurate beliefs?

One promising approach to limiting the effects of misinformation was already on the rise before the COVID-19 pandemic: increasing misinformation resistance through educational interventions. A substantial number of countries have implemented educational interventions, primarily focused on *media literacy* [10], which can be understood as the ability to access, analyze, evaluate, and communicate messages in a variety of forms [11]. The Swedish Civil Contingencies Agency, for instance, has included a section about misinformation in its public emergency preparedness brochure, advising Swedes to be aware of the aim of information and to check the source of information, among others [12]. Similarly, Facebook tries to help its users recognize misinformation by providing 10 tips [13]. One advantage of such a focus on media literacy is that it can help prevent problems with misinformation, instead of having to correct false beliefs after they have taken hold. Previous media literacy research, with interventions focusing on identification of misinformation, has yielded promising results indicating that some interventions can reduce the perceived accuracy of misinformation [14,15]. Other research highlights the difficulties in crafting media literacy interventions [16]. Can these types of interventions, focusing on empowerment of media consumers, help individuals deal with the supposed COVID-19 infodemic?

Our approach focuses on helping individuals figure out what is true and what is false, considering false such beliefs about factual matters that are not supported by clear evidence and expert opinion [17]. We test an intervention that empowers people to search for and identify scientific consensus. Communicating scientific consensus (ie, a high degree of agreement between scientists) is effective in eliciting scientifically accurate beliefs [18]. This effectiveness is described in the gateway belief model, which states that people's perceived scientific consensus functions as a gateway to their personal factual beliefs [19,20]. Here, we focus on empowering individuals to search for and identify scientific consensus, because this approach is more flexible than communicating a scientific consensus on every single issue.

The current strategy is considered a *boosting* approach. Boosting encompasses interventions targeting competence rather than immediate behavior [21]. In line with this, our intervention focuses on improving people's skills to form accurate beliefs, instead of altering the external context within which people form beliefs. In addition, our boosting approach can be considered an educational intervention, just like media literacy interventions. However, compared to media literacy interventions that target the identification of misinformation, boosting *consensus reasoning* is not dependent on being exposed to misinformation. One can investigate any claim, true or false, from any source.

This study had two main goals. One involved an exploratory, not preregistered, investigation to gain insight into the effects of the supposed infodemic on individuals' belief accuracy in times of crisis, and to investigate potential correlates of belief accuracy. The second goal was a preregistered test of the boosting intervention aimed at increasing belief accuracy. Accordingly, we hypothesized that our intervention would lead to more accurate beliefs about the COVID-19 pandemic than the control condition; the complete preregistration can be found on the Open Science Framework (OSF) [22]. The research was conducted online, and a balanced sample of the US population was recruited. We decided to focus on the United States, because this is arguably the country worst hit by the COVID-19 pandemic. Using a longitudinal design, measuring beliefs about the pandemic over 4 weeks just after daily confirmed COVID-19 deaths had peaked at over 4000, allowed us to investigate belief formation in the relatively early days of the pandemic. All data and material are available on the project page on the OSF [23].

Methods

Recruitment

We used Prolific, a UK-based online crowdsourcing platform that connects researchers to participants, to collect data from US citizens over a 4-week period. Prolific has been demonstrated to yield high-quality data and more diverse participants than student samples or other major crowdsourcing platforms [24]. In addition, it allowed for recruitment that balanced age, gender, and ethnicity to approximate the US general public, via stratification using US census data [25]. Recruitment for the initial baseline wave started on April 27, 2020.

A total of 1212 individuals participated in the study at baseline (T0), for which they received £0.45 (~US $0.56). A total of 1089 individuals participated in the first follow-up wave (T1), 1070 individuals participated in the second follow-up wave (T2), and 1028 individuals participated in the final wave (T3); see Table 1 for total sample size, exclusions, and final sample size per wave. Participants received £0.33 (~US $0.41) for participation per follow-up survey. Each of the waves was separated by approximately one week (mean$_{\text{T0-T1}}$ 6.98 days, SD $_{\text{T0-T1}}$ 14.92 hours; mean$_{\text{T1-T2}}$ 7.01 days, SD $_{\text{T1-T2}}$ 12.72 hours; mean$_{\text{T2-T3}}$ 7.06 days, SD $_{\text{T2-T3}}$ 13.18 hours). The sample size was determined by the available resources. This study is part of a research project that was reviewed and approved by the Ethics Committee Social Science at Radboud University (reference No. ECSW-2018-056).

Table 1. Total sample size, exclusions, and final sample size per wave.

Wave and sample	Value, n (%)	Retention rate, %[a]
T0 (baseline): April 27-29, 2020		
Total sample	1212 (100)	N/A[b]
Excluded[c]	10 (0.8)	N/A
Final sample	1202 (99.2)	N/A
T1 (first follow-up wave): May 4-7, 2020		
Total sample	1089 (100)	N/A
Excluded[c]	11 (1.0)	N/A
Final sample	1078 (99.0)	89.7
T2 (second follow-up wave): May 11-14, 2020		
Total sample	1070 (100)	N/A
Excluded[c]	3 (0.3)	N/A
Final sample	1067 (99.7)	88.8
T3 (final wave): May 18-21, 2020		
Total sample	1028 (100)	N/A
Excluded[c]	6 (0.6)	N/A
Final sample	1022 (99.4)	85.0

[a] The retention rate is based on the final sample sizes of T0 and the respective wave. Total sample sizes of follow-up waves were counted, excluding 2 participants who should have been excluded but had been allowed to participate in the follow-up waves due to a technical error.

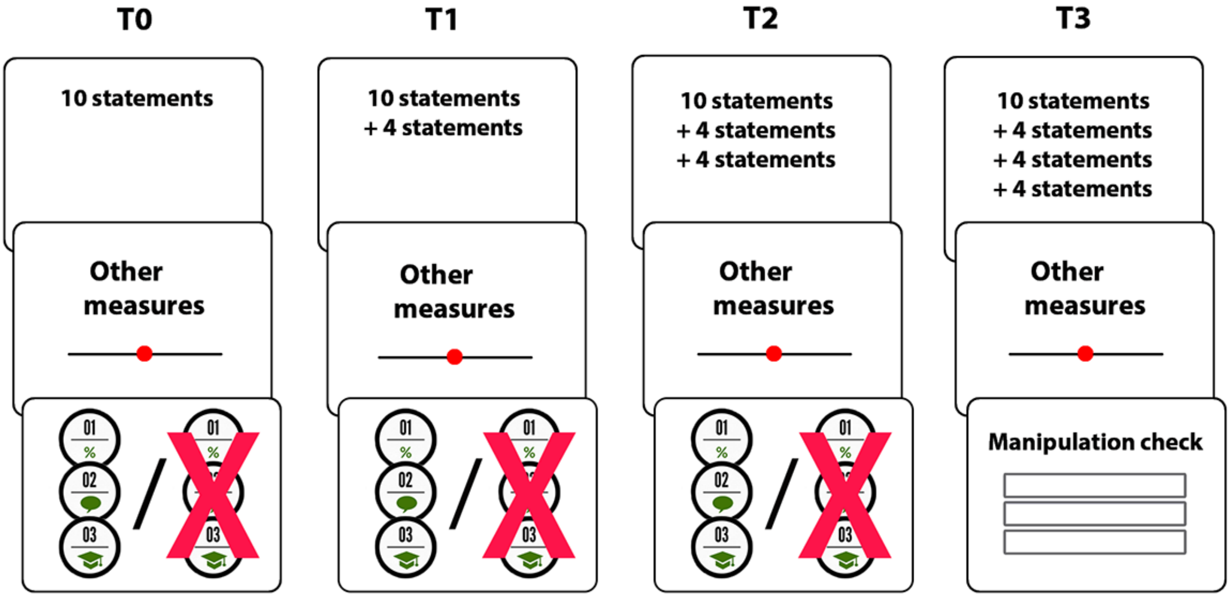
[b] N/A: not applicable; retention rates were calculated using final sample sizes of T0 and the follow-up waves.

[c] More details on exclusions can be found in the Statistical Analysis, Data Exclusion section.

Procedure

Participants were randomly assigned to either receive the intervention (ie, the boost condition) or no intervention (ie, the control condition). All surveys started with the measure of belief accuracy, for which participants were presented with 10 (T0), 14 (T1), 18 (T2), or 22 (T3) statements about the coronavirus and COVID-19 (see Figure 1). Participants indicated to what extent they believed each statement to be true. An attention check was included among these statements (see Multimedia Appendix 1). Subsequently, participants reported their behavior aimed at preventing the spread of the coronavirus and completed the other measures. The boosting intervention, an infographic presenting three steps that can be used to evaluate a claim, was included at the end of T0, T1, and T2. Only participants in the boost condition were presented with the infographic, allowing them to apply their boosted consensus reasoning skill in the week leading up to the next wave. At T3, all participants completed a manipulation check. At the end of T0, all participants entered demographic information and completed a seriousness check (see Multimedia Appendix 1). All surveys took about 3 to 6 minutes.

Figure 1. Flowchart of the main elements of the procedure per wave. Participants first completed the measure of belief accuracy, then completed other measures, and, finally, were exposed to the intervention or not. At T3, participants completed a manipulation check. The bottom panels of the first three columns display the intervention condition (left; infographic) and the control condition (right; no intervention). T0: baseline; T1: first follow-up wave; T2: second follow-up wave; T3: final wave.



Materials and Measures

Belief Accuracy

The key dependent variable was the accuracy of participants' beliefs related to the COVID-19 pandemic. This variable consisted of responses to a number of statements about the pandemic, which were sourced from preprints of early research on public perceptions of COVID-19 (eg, [26]), public health agencies and medical institutes (eg, the WHO), media tracking organizations (eg, NewsGuard), and expert reports in established media (eg, CNBC); a comprehensive list of these resources is available in Multimedia Appendix 2. Only statements based on scientific claims were included in order to make sure that there was compelling evidence that the claims were either true or false.

At T0, participants were exposed to 10 statements, of which five were scientifically accurate (eg, "Fever is one of the symptoms of COVID-19") and five were at odds with the best available evidence (eg, "Radiation from 5G cell towers is helping spread the coronavirus"). Participants responded by indicating the accuracy of a statement as follows: *false*, *probably false*, *don't know*, *probably true*, or *true*. In each subsequent wave, four new statements were added to the list of statements: two accurate ones and two inaccurate ones. This allowed us to keep the belief accuracy measure current, reflecting contemporary insights and discussion points. The order of the statements was randomized per participant and varied per wave.

A belief accuracy score was calculated by converting the response to each statement to a number reflecting how accurate the response was; a correct judgment was counted as 1 and an incorrect judgment was counted as –1. A less certain but correct *probably true* or *probably false* counted as 0.5 and an incorrect one as –0.5. Finally, a *don't know* response was counted as 0. Average scores were calculated per wave per participant, resulting in a repeated measure of belief accuracy. Internal consistency was acceptable to good across the four waves; the McDonald ω_t was between 0.75 and 0.87 in all waves.

Coronavirus-Related Behavior

Coronavirus-related behavior aimed at preventing the coronavirus from spreading was measured by asking participants to indicate their agreement with three statements. The statements were "To prevent the coronavirus from spreading..." (1) "I wash my hands frequently," (2) "I try to stay at home/limit the times I go out," and (3) "I practice social distancing (also referred to as 'physical distancing') in case I go out"; agreement was measured on a scale from 1 (strongly disagree) to 7 (strongly agree). Scores were averaged per wave per participant. Internal consistency was acceptable to good across the four waves; the McDonald ω_t was between 0.77 and 0.83 in all waves.

Additional Measures

Trust in scientists was measured in all four waves with responses to the statement "I trust scientists as a source of information about the coronavirus." Participants responded on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Participants' primary news source for information about the COVID-19 pandemic was identified by asking them at T0 what their main source of news about the coronavirus was. Participants could choose one option from a list of 11 news sources, based on data from the Pew Research Center on Americans' news habits [27].

Finally, we included a manipulation check at T3. This consisted of asking participants how they evaluated the truthfulness of the statements about the coronavirus and coronavirus disease in the study over the past weeks. We asked them to name the steps that they took to evaluate the claims in three open text boxes, of which at least one had to be used. These answers were coded by the first author to indicate whether they mention consensus—or something similar—or not. A second coder coded a random subset of 120 answers, with Krippendorff α indicating good (α=.85) interrater reliability. Therefore, the complete coding from the first author was used in the analyses.

Not all measures included in the study are listed here, because not all measures are relevant here. Please see the material on the project page on the OSF for the remaining measures [23].

Intervention

The boosting intervention that was included at the end of T0, T1, and T2 consisted of a short infographic that was aimed at empowering participants to use scientific consensus when evaluating claims related to the COVID-19 pandemic. The infographic set out three steps that can be used to evaluate a claim: (1) searching for a statement indicating consensus among scientists, (2) checking the source of this consensus statement, and (3) evaluating the expertise of the consensus. The infographic can be found in Multimedia Appendix 3. Participants in the control condition were not exposed to the infographic.

Demographics

Demographics including political orientation, age, gender, ethnicity, and education were asked about at T0. Political orientation was measured by combining political identity (ie, strong Democrat, Democrat, independent lean Democrat, independent, independent lean Republican, Republican, or strong Republican) and political ideology (ie, very liberal, liberal, moderate, conservative, or very conservative) into one numeric, standardized measure centered on 0 (ie, moderate, independent), based on Kahan [28].

Statistical Analysis

Data Exclusion

First, we removed one of two duplicate responses at T1 and excluded all responses from one participant with three varying responses at T3.

As preregistered, participants who failed the attention check at T0 were excluded and replaced (n=8; including 2 who had been allowed to participate in the follow-up waves due to a technical error). If a participant failed one of the attention checks in the subsequent waves, data from that wave was not included in the analyses (n_{T1}=5, n_{T2}=2, n_{T3}=5), but other surveys in which the attention check was passed were retained. Participants who indicated at the T0 seriousness check that their data should not be used were excluded from further participation and their data were not used, but they were not replaced (n=2). No participants completed T0 in less than 1 minute, but if a participant completed a subsequent wave in less than 1 minute, data from that survey were not included in the analyses (n_{T1}=6, n_{T2}=1, n_{T3}=0). Other waves in which the 1-minute threshold was passed were retained.

Exploratory Analyses

General increase in belief accuracy over time was explored using linear mixed modeling for each set of statements, with wave as predictor, controlling for political orientation and including a random intercept per participant. The relationship between belief accuracy and coronavirus-related behavior was explored with correlations for each wave. The relationship of belief accuracy with trust in scientists at T0, political orientation, and primary news source was explored using mixed modeling, controlling for wave, age, gender, education, and ethnicity. The interaction term between trust and political orientation was included in the model. The five most chosen news sources (ie, CNN, Fox News, NPR, social media sites, and The New York Times, excluding the option *Other sources*) were included as dummy-coded variables. Finally, we included a random intercept and a random slope for wave per participant. Mixed modeling was performed with the lme4 package [29] in R (The R Foundation) [30]. The models were examined using likelihood ratio tests, using the R package lmerTest [31].

Preregistered Analysis

The hypothesis that our intervention would lead to more accurate beliefs than would the control condition was also tested using linear mixed modeling. The condition (ie, intervention vs control) and wave, and the interaction between condition and wave, were included as predictors in the model. Political orientation was included as a covariate, because beliefs about the COVID-19 pandemic are related to political ideology [32], and a random intercept and a random slope for wave were included per participant. The hypothesis was tested by comparing the full model, with the interaction between condition and wave, to a model without this interaction effect. We used the PBmodcomp function from the R package pbkrtest [33] for parametric bootstrapping (10,000 simulations).

Results

Participants

The final sample roughly reflects US census data [25] on gender, age, and ethnicity, indicating that the balanced sampling worked well. See Table 2 for more details.

Table 2. Participant characteristics.

Demographic characteristic	Sample value (N=1202), n (%)[a]	Census value, %[b]
Gender		
Female	604 (50.2)	51.3
Male	587 (48.8)	48.7
Other	11 (0.9)	N/A[c]
Age in years		
18-24	164 (13.6)	11.9
25-34	243 (20.2)	17.9
35-44	209 (17.4)	16.4
45-54	199 (16.6)	16.0
55-64	232 (19.3)	16.6
65-74	139 (11.6)	12.4
≥75	16 (1.3)	8.8
Ethnicity		
White	918 (76.4)	73.6
Black	158 (13.1)	12.5
Asian	79 (6.6)	5.9
Mixed	30 (2.5)	2.5
Other	17 (1.4)	5.5

[a] Due to rounding, percentages may not add up to 100% exactly.

[b] The percentages in the census data reflect the population aged 18 years and over.

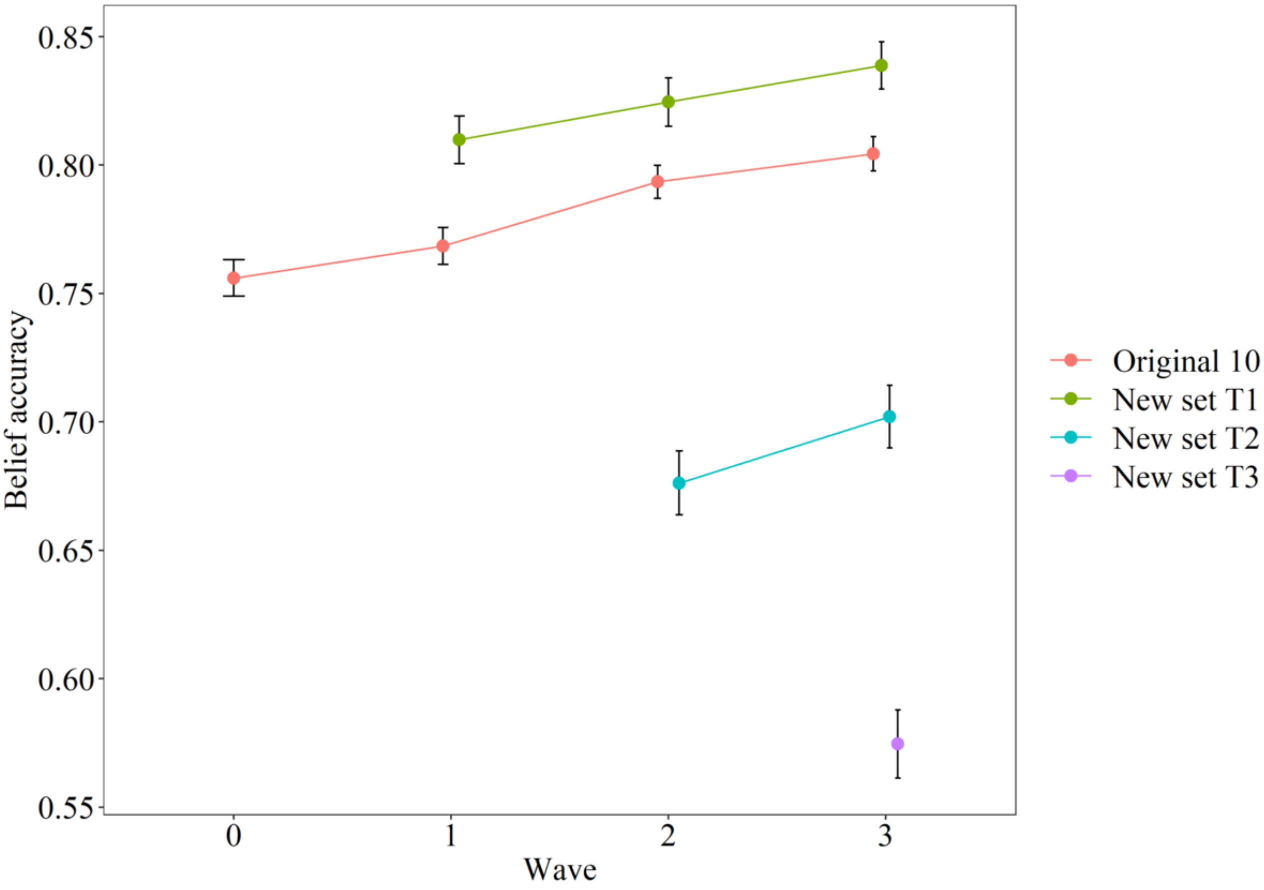
[c] N/A: not applicable.

Belief Accuracy

Mean scores of belief accuracy were very high for all waves, with scores reflecting low belief in false statements and high belief in true statements. There was substantial variation in the accuracy of responses between statements, although none of the statements was ever interpreted with less than 0.25 accuracy, on average; see Multimedia Appendix 4 for a complete overview of scores per statement per wave.

There was a modest increase in belief accuracy over time, looking at each set of statements separately (first 10: estimate=0.02, SE<0.01; $t_{3202.59}$=13.82, *P*<.001; T1 set: estimate=0.01, SE<0.01; $t_{2041.94}$=4.80, *P*<.001; T2 set: estimate=0.02, SE<0.01; $t_{1003.22}$=3.40, *P*<.001). This increase was positive for all three sets of statements that were asked more than once (see Figure 2 [34,35]), indicating that participants became more accurate in their interpretation of the statements over time.

Figure 2. Belief accuracy per set of statements over time. The new set at T3 was included for completeness. Focusing on within-subject change, dots represent normed means and error bars indicate 95% CIs of the within-subject SE [34], calculated using the summarySEwithin function from the Rmisc package [35]. The belief accuracy scale ranged from –1, indicating completely inaccurate beliefs, to 1, indicating completely accurate beliefs. T1: first follow-up wave; T2: second follow-up wave; T3: final wave.



Coronavirus-Related Behavior

Accurate beliefs were correlated with self-reported behavior aimed at preventing the coronavirus from spreading; *r* at all waves was between 0.26 and 0.29 and all *P* values were less than .001. This small but robust correlation suggests that accurate beliefs could be important for coronavirus-related behavior. We explored potential evidence of any causal effects in the data using a random-intercept cross-lagged panel model. This yielded a tentative indication that accurate beliefs might be predictive of behavior, with belief accuracy at T2 predicting coronavirus-related behavior at T3. However, with all other paths showing no sign of significant predictive effects, the results regarding causality are largely inconclusive (see Multimedia Appendix 5).

Associations With Belief Accuracy

We explored the relationship of trust in scientists at T0, political orientation, and the primary news source with belief accuracy. The mixed model yielded a significant positive relationship between belief accuracy and trust (estimate=0.07, SE<0.01; $t_{1200.23}$=16.44, *P*<.001) and a significant negative correlation with political orientation (estimate=–0.02, SE<0.01; $t_{1199.62}$=–6.78, *P*<.001). These main effects indicated that participants with higher trust in scientists scored higher on the measure of belief accuracy and that liberal, Democratic participants held more accurate beliefs than conservative, Republican participants. Moreover, these main effects were partially qualified by an interaction effect among trust and political orientation (estimate=–0.01, SE<0.01; $t_{1195.05}$=–3.62, *P*<.001). Plotting of this interaction effect demonstrated that trust in scientists had a stronger relationship with belief accuracy for liberal, Democratic participants than it had for conservative, Republican participants (see Figure 3).

Two of the five most chosen primary news sources were associated with a worse understanding of the facts regarding the COVID-19 pandemic than others (see Figure 4). Participants who reported CNN (estimate=–0.03, SE=0.01; $t_{1194.49}$=–2.33, *P*=.02) or Fox News (estimate=–0.05, SE=0.02; $t_{1202.49}$=–3.05, *P*=.002) as their main news source scored below average on belief accuracy.

Figure 3. Linear relationship between belief accuracy (averaged over wave for plotting) and trust in scientists at T0 (baseline), split by political orientation (dichotomized for plotting). The grey area represents the 95% CI. The belief accuracy scale ranged from –1, indicating completely inaccurate beliefs, to 1, indicating completely accurate beliefs. Trust in scientists was measured with responses to the statement “I trust scientists as a source of information about the coronavirus.” Participants responded on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree).

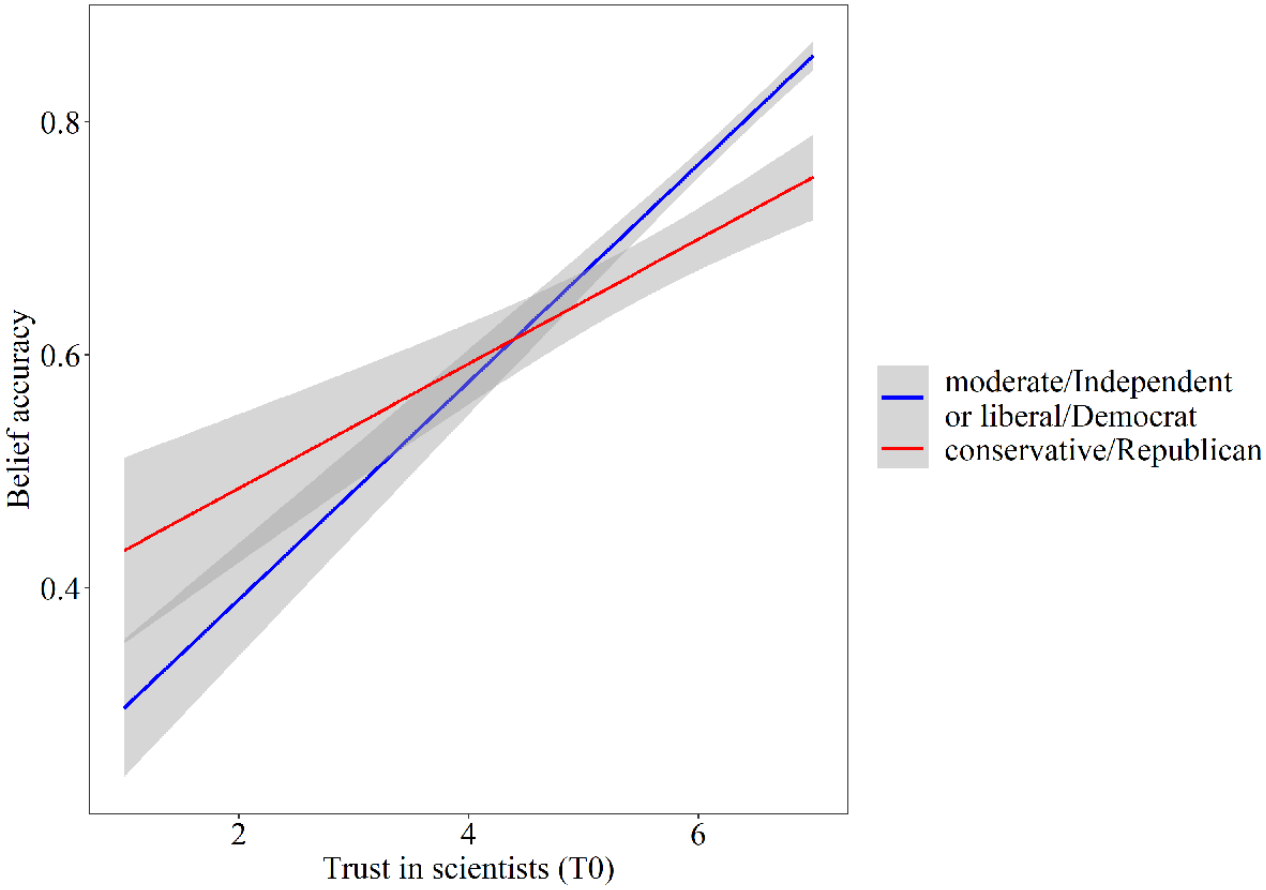
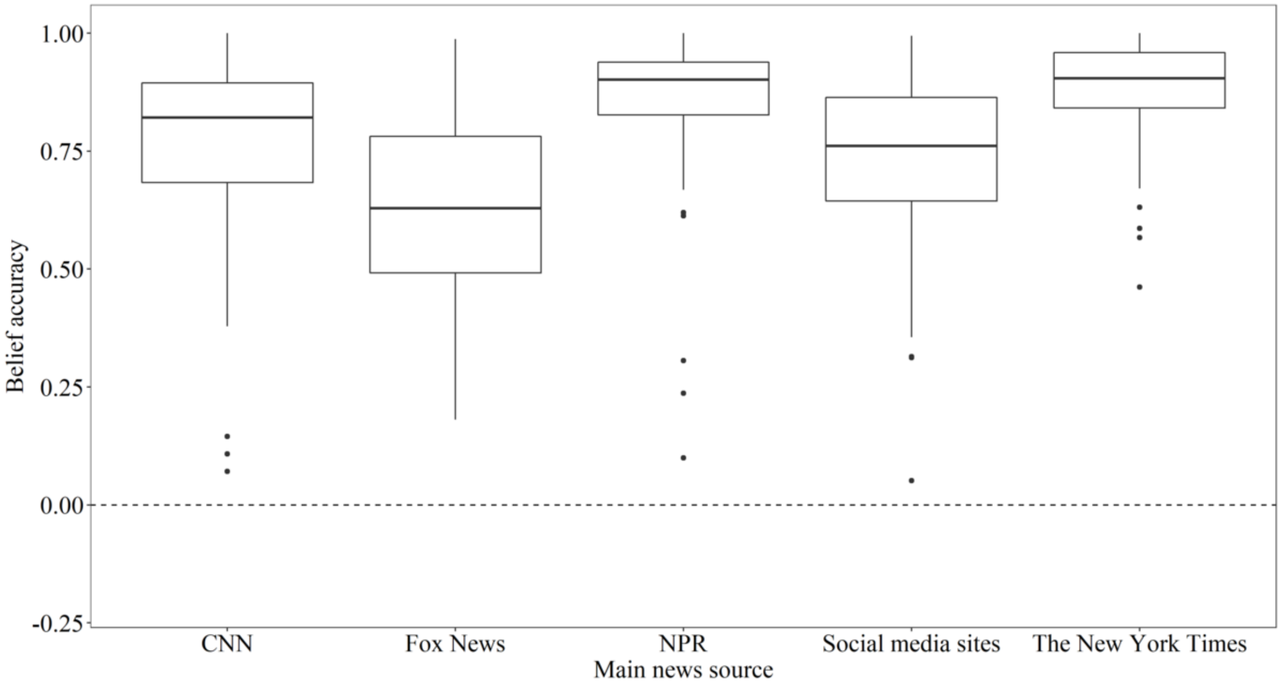


Figure 4. Box plot of raw (unadjusted) scores of belief accuracy (averaged over wave for plotting) by main news source. The belief accuracy scale ranged from –1, indicating completely inaccurate beliefs, to 1, indicating completely accurate beliefs.

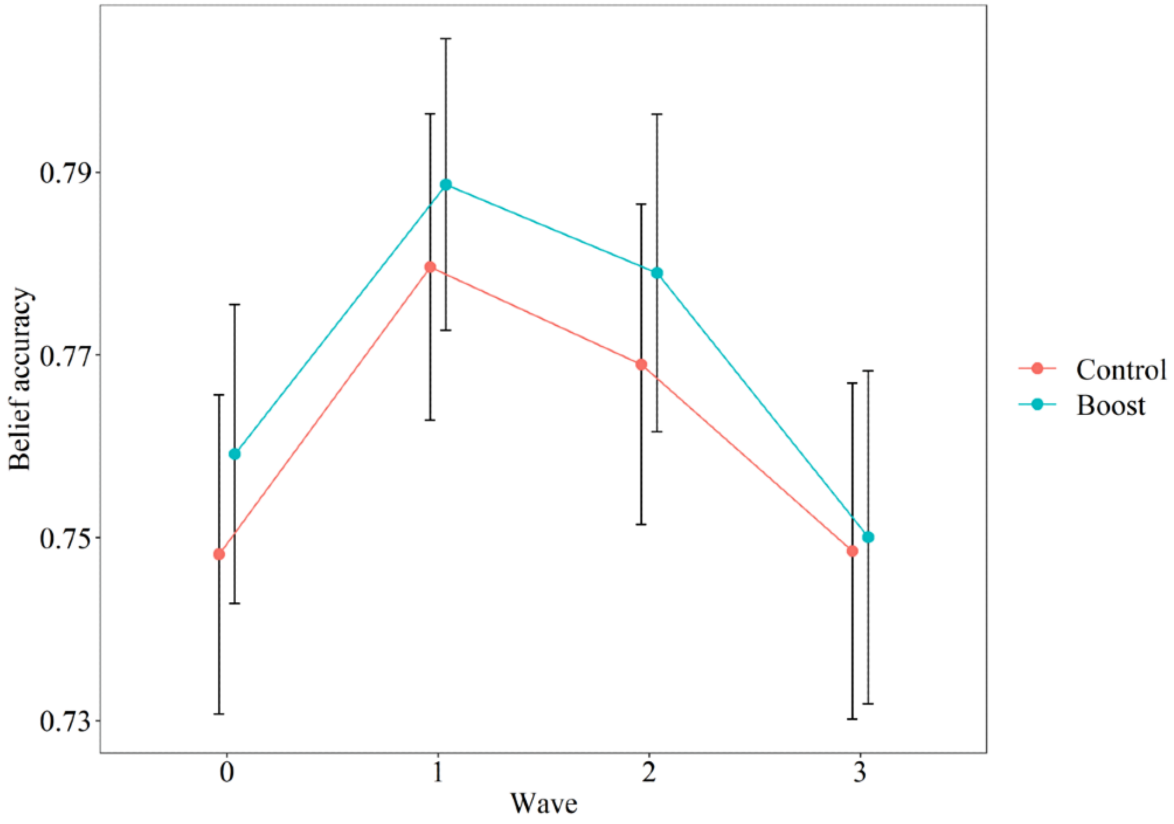


Intervention

We conducted a manipulation check and, as expected, when asked how they evaluated claims, participants in the boost condition (136/600, 22.7%) mentioned consensus, or something similar, more often than participants in the control condition (26/602, 4.3%; χ^2_1=85.2, *P*<.001, N=1202).

We hypothesized that our boosting intervention would lead to more accurate beliefs about the COVID-19 pandemic than would the control condition. However, the interaction effect between condition and wave on belief accuracy was not significant (estimate<0.01, SE<0.01; $t_{1074.36}$=0.22, *P*=.83). This means that the boosting intervention did not significantly alter belief accuracy of participants over time, compared to the control condition (see Figure 5). This was also the case when we explored effects of the intervention on inaccurate statements only (*P*=.48), accurate statements only (*P*=.49), only the original 10 statements that were included in all waves (*P*=.61), and only included participants who scored relatively low on belief accuracy at T0 (belief accuracy_{T0}<0.76; *P*=.32).

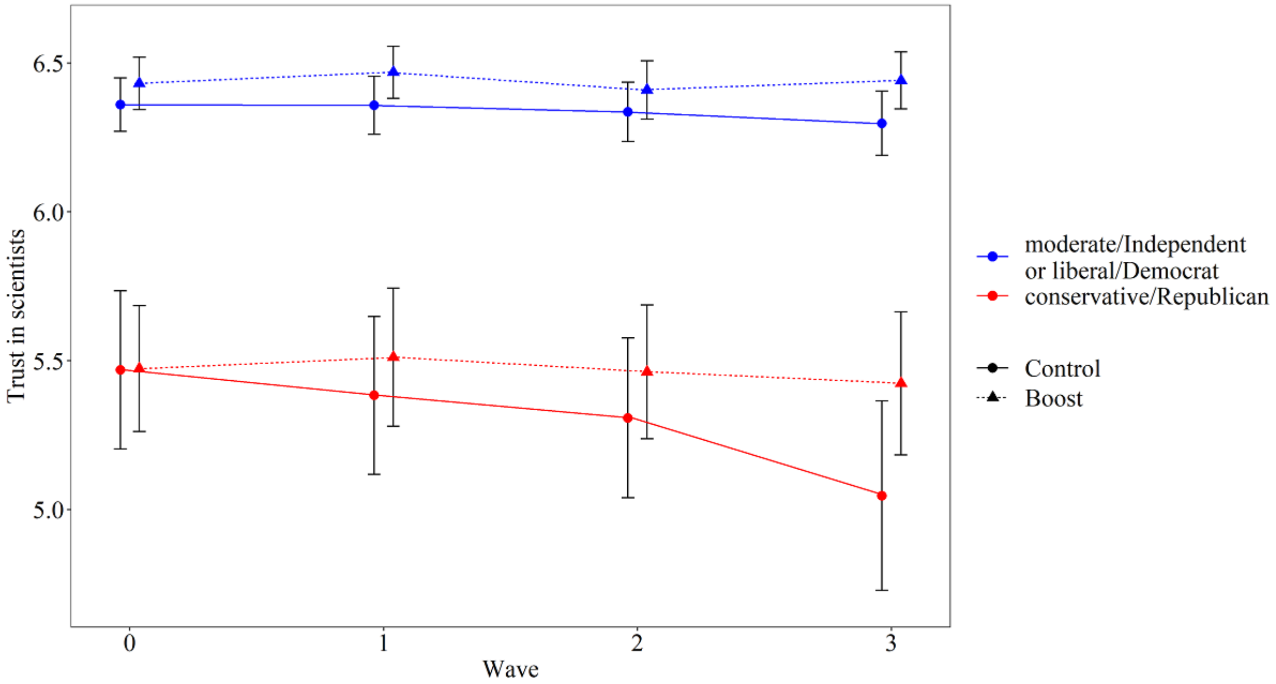
Figure 5. Belief accuracy per condition over time. Error bars indicate 95% CI focusing on the comparison between experimental conditions, not adjusted for within-subject variability. The belief accuracy scale ranged from –1, indicating completely inaccurate beliefs, to 1, indicating completely accurate beliefs.



When the subsample of participants in the boost condition who mentioned consensus, or similar, in the manipulation check was compared to participants in the control condition, we again found that the intervention did not increase belief accuracy (*P*=.21). However, there was a main effect of condition (estimate=0.06, SE=0.02; $t_{725.66}$=3.50, *P*=.001), indicating that participants in the boost condition who did use consensus while evaluating claims scored higher on belief accuracy than participants in the control condition. This difference was already present at T0, so was not caused by the intervention.

We explored the effect of the boosting intervention on trust in scientists as a source of information about the coronavirus. The mixed-effects model, similar to the hypothesis test but with the repeated measure of trust as the dependent variable and including the interaction term between condition, wave, and political orientation, yielded a significant three-way interaction effect between condition, wave, and political orientation (estimate=0.02, SE=0.01; $t_{1088.42}$=2.39, *P*=.02). Trust in scientists was very high in all four waves (means between 6.11 and 6.19), but investigation of the two-way interaction effects per condition indicated a significant interaction effect among wave and political orientation in the control condition (estimate=–0.02, SE=0.01; $t_{562.01}$=–3.24, *P*=.001), while there was no such significant interaction effect in the boost condition (*P*=.90). As illustrated by Figure 6, there was a clear overall difference in trust in scientists between participants related to their political orientation. More interestingly, trust remained stable for all participants in the boost condition, but decreased slightly for conservative, Republican participants in the control condition. This could indicate that the boosting intervention inhibited a decline of trust in scientists as a source of information about the coronavirus among more conservative, Republican participants.

Figure 6. Trust in scientists as a source of information about the coronavirus per condition and political orientation (dichotomized for plotting) over time. Error bars indicate 95% CI focusing on the comparison between experimental conditions, not adjusted for within-subject variability. Trust in scientists was measured with responses to the statement "I trust scientists as a source of information about the coronavirus." Participants responded on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree).



Discussion

Principal Findings

The aims of this study were to gain insight into the beliefs of the US public about the COVID-19 pandemic and to investigate whether a boosting intervention could improve people's belief accuracy. Interestingly, the average scores on belief accuracy over the surveyed 4-week period were high, indicating low belief in false statements and high belief in true statements. Looking at each set of statements, we found a small but significant increase in belief accuracy over time. This indicates that the general public is quite able to figure out what is true and what is not in times of crisis. Moreover, a small but robust correlation suggests that accurate beliefs about the pandemic might be important for coronavirus-related behavior. Associations with belief accuracy suggest that the processes of belief formation and correction might be affected by individuals' trust in scientists and political orientation, as well as their news habits. Finally, the boosting intervention yielded no significant increase in belief accuracy over the control condition, demonstrating that the boosting infographic was not successful in helping people figure out what is true and what is false. Exploratory analyses suggested that the intervention did, however, inhibit a decline in trust in scientists as a source of information about the coronavirus among more conservative, Republican participants.

Comparison With Prior Work

There is a great deal of worry about the prevalence of misinformation during the current pandemic, which is reflected in popular media (eg, [36,37]), as well as among scientists (eg, [38,39]) and public health agencies (eg, [4,40]). The supposed COVID-19 infodemic is not reflected in US citizens' beliefs. The finding that most Americans hold quite accurate beliefs about the COVID-19 pandemic is in line with emerging work on perceptions of the pandemic that shows that belief in COVID-19 misperceptions and conspiracy theories is quite low [41-44]. Consequently, this calls into question the prevalence of misinformation or the public's susceptibility to misinformation.

A convincing body of empirical work on the prevalence of misinformation surrounding the COVID-19 pandemic is not yet available. Research from before the COVID-19 pandemic indicates that the prevalence of misinformation might be lower than many believe [45-47]. Still, it is possible that the current pandemic has led to an increase in misinformation compared to the information landscape from before the pandemic. However, when looking for potential explanations of this study's findings, we should consider the possibility that COVID-19 misinformation is not as prevalent as expected. Perhaps misinformation makes up only a small portion of the average US citizen's media diet.

The second possibility is that we are indeed facing a COVID-19 infodemic, but that the public is not very susceptible to it. Misinformation campaigns regarding other topics, such as climate change and the health effects of tobacco [18,48], have demonstrated that misinformation can contribute to misperceptions about important matters. In these cases, however, misinformation campaigns have been carefully organized and executed, continually misinforming the public for decades. In contrast, the COVID-19 pandemic is a novel issue and, at least in the relatively early months that we investigated, did not yield many such coordinated misinformation campaigns. Moreover, the COVID-19 pandemic originated in a very different media landscape than the climate change and tobacco misinformation campaigns. Fake news, misinformation, and disinformation have been discussed widely and frequently in popular media since the 2016 US presidential election and the 2016 United Kingdom European Union membership referendum. This might have resulted in the public being more aware of misinformation campaigns targeting them. Perhaps the widespread discussion of misinformation in popular media has worked as a large-scale media literacy intervention, putting people *on guard* against false information. In support of this idea, recent research has demonstrated that simply asking one to consider the accuracy of a claim improved subsequent choices about what COVID-19 news to share on social media [49].

A third possibility that should be considered is that the public is more careful in forming beliefs in times of crisis, especially in the relatively early days of a crisis, making a well-informed public not unique to the COVID-19 pandemic. In times of crisis, people are likely to increase news consumption (eg, [50]). This was also the case in the United States during the first months of the COVID-19 pandemic, with people reporting increased news consumption [51]. Although we found that those who reported CNN or Fox News as their main news source scored below average on belief accuracy, the general increase in news consumption may lead to a better understanding of the crisis situation, including more accurate beliefs.

Turning to the finding that political orientation is associated with individuals' belief accuracy, we see that this is in line with other emerging work [41]. There is likely a multitude of explanations for this evolving partisan divide [52] regarding perceptions about the pandemic, such as political party cues in the news affecting opinion formation [53], the difficulty of correcting false beliefs for the ideological group most likely to hold those misperceptions [17], as well as the differences in news consumption that are reflected in this study. A second variable that is even more strongly related to belief accuracy is trust in scientists as a source of information about the coronavirus, demonstrating that higher trust is related to more accurate beliefs. Interestingly, the associations of political orientation and trust with accurate beliefs were partially explained by an interaction effect among political orientation and trust. The stronger association of trust with belief accuracy for more liberal, Democratic individuals might mean that they rely more on scientists' perceptions in forming beliefs, while relatively more conservative, Republican individuals might rely more on other cues. Relatedly, the inhibited decline in trust in scientists among conservative, Republican participants in the boost condition could indicate that information about the scientific process might resonate more with them than just hearing the results of this process. Though this exploratory finding should be replicated, it could provide a fruitful avenue for further research on trust in scientists and political orientation.

In addition, this study demonstrates that some news sources might be doing a worse job of informing their consumers about the COVID-19 pandemic than others, or perhaps that better-informed news consumers turn to different news sources than less well-informed consumers, again in line with other emerging work [41]. Most likely, a combination of both selection and influence (eg, [54]) explain the differences in belief accuracy found in this study. Interestingly, considering the role of social media in the spread of misinformation (eg, [55]), with about 26% to 42% of tweets in the data collection period containing unreliable facts [56], participants who reported social media sites as their main source of news about the coronavirus did not display significantly worse belief accuracy than others. However, it is possible that participants who reported social media sites as their main source followed major news outlets via the social media site, thereby being exposed to similar news content as the other participants.

Finally, this study demonstrates the difficulty of crafting interventions aimed at increasing belief accuracy. Recent work demonstrates that simple, short media literacy interventions can work [14,15], while other work highlights the difficulties of crafting these interventions [16]. We argue that the divergent findings can be explained by the fact that in the former work the interventions were paired with corrections, while in this study participants had to put their new skill to use outside of the study context. Considering that cues signaling the existence of consensus in relevant news content are very rare [57], participants likely had to search for information about scientific consensus themselves. The results from the manipulation check indicated that only a relatively small portion of participants actually applied this strategy. However, those individuals who indicated that they did apply a strategy related to consensus reasoning scored higher on belief accuracy than the control group. This difference highlights the potential of the intervention in situations where individuals can be empowered to actually apply it.

Limitations

There are two notable limitations to this study. First, our belief accuracy measure consisted only of science-based statements. We incorporated only science-based claims in our study to ensure that there was sufficient empirical evidence stating that a claim was either true or false. However, this decision did exclude some coronavirus-related claims that were not based on science (eg, "Bill Gates patented the coronavirus") or were unresolved at the time (eg, "A vaccine will be available before the end of the year"). It should have been harder for participants to figure out whether such unresolved issues were true or not, yielding different responses from participants for a measure reflecting non-science-based, unresolved issues about the pandemic.

A second limitation is the fact that the recruitment platform that we used, Prolific, is known as a platform for research. Although participants on the platform receive financial incentives for completing studies, they might be more interested in scientific research than the average US citizen. This could lead to them also having a higher trust in science than the general population, even though our sample was balanced regarding age, gender, and ethnicity. As trust in science was highly related to belief accuracy, it could be possible that this led to an inflated belief accuracy score. Future research should attempt to replicate this study with a sample that represents the US population better than our balanced sample.

Conclusions

Our work demonstrates that most people are quite able to figure out the facts in this time of crisis, but also that it is difficult to craft an intervention targeting these beliefs. However, in cases where people do not immediately have a clear understanding of the facts, they are capable of figuring them out over time. There are some factors that might make it easier or harder for one to figure out the facts. We found that the accuracy of participants' beliefs was related to political orientation, as well as their primary news source. This suggests that, even in the relatively early days of the pandemic, political polarization and media diet had a grip on US citizens' factual beliefs, leading to polarization along party lines. Another factor strongly related to accurate beliefs about the pandemic was trust in scientists. It is unclear whether an already-high trust led to accurate beliefs or whether being able to figure out the facts increased trust in scientists, but the importance of expert communication is underlined by these findings.

Although a small but robust correlation suggests that accurate beliefs about the pandemic might be important for coronavirus-related behavior, the role of misinformation in the pandemic seems to be relatively small, either because it is rare or because it is unable to persuade. However, we note that even if misinformation is not prevalent and only accepted by a small portion of the receivers, it can still be dangerous. To illustrate, we found that almost all participants in this study disregarded the statement that injecting or ingesting bleach is a safe way to kill the coronavirus, but this false claim is reported to have cost at least one life [58]. Additionally, with the antivaccine community launching coordinated misinformation campaigns against potential coronavirus vaccines [59] and politicization of the pandemic looming, the infodemic might become a much bigger threat.

Acknowledgments

The research was funded solely by the first author's research institute and supported by Prolific's generous decision to waive service fees on COVID-19-related research.

Authors' Contributions

AvS designed the experiments, conducted the experiments, and analyzed the data. GS, HV, and MB advised on the design of the experiments and analyses. All authors contributed to writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Attention and seriousness checks.
[DOCX File , 17 KB - jmir_v23i1e24069_app1.docx]

Multimedia Appendix 2
Resources used to collect COVID-19 pandemic belief statements.
[DOCX File , 18 KB - jmir_v23i1e24069_app2.docx]

Multimedia Appendix 3
Infographic used in the boosting intervention to empower participants to use the scientific consensus when evaluating claims related to the COVID-19 pandemic.
[PNG File , 779 KB - jmir_v23i1e24069_app3.png]

Multimedia Appendix 4
Descriptive statistics of all statements per wave.
[DOCX File , 20 KB - jmir_v23i1e24069_app4.docx]

Multimedia Appendix 5
Random-intercept cross-lagged panel model (RI-CLPM).
[DOCX File , 20 KB - jmir_v23i1e24069_app5.docx]

References

1. Rahim Z. More than 500,000 people have been killed by Covid-19. A quarter of them are Americans. CNN. 2020 Jun 29. URL: https://edition.cnn.com/2020/06/29/world/coronavirus-death-toll-cases-intl/index.html [accessed 2021-01-05]

2. Mitchell A, Baxter Oliphant J, Shearer E. Majority of Americans feel they need breaks from COVID 19 news; many say it takes an emotional toll. Pew Research Center. Washington, DC: Pew Research Center; 2020 Apr 29. URL: https://tinyurl.com/y2gwaek6 [accessed 2021-01-05]
3. Frenkel S, Alba D, Zhong R. Surge of virus misinformation stumps Facebook and Twitter. The New York Times. 2020 Mar 08. URL: https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html [accessed 2021-01-05]
4. United Nations Department of Global Communications. UN tackles ‘infodemic’ of misinformation and cybercrime in COVID-19 crisis. United Nations. 2020 Mar 31. URL: https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19 [accessed 2021-01-05]
5. Romm T. Facebook will remove misinformation about coronavirus. The Washington Post. 2020 Jan 31. URL: https://www.washingtonpost.com/technology/2020/01/30/facebook-coronavirus-fakes/ [accessed 2021-01-05]
6. Allyn B. Twitter now labels 'potentially harmful' coronavirus tweets. NPR. 2020 May 11. URL: https://www.npr.org/sections/coronavirus-live-updates/2020/05/11/853886052/twitter-to-label-potentially-harmful-coronavirus-tweets [accessed 2021-01-05]
7. Facebook's algorithm: A major threat to public health. Avaaz. 2020 Aug 19. URL: https://avaazimages.avaaz.org/facebook_threat_health.pdf [accessed 2021-01-05]
8. Milman O. Trump touts hydroxychloroquine as a cure for Covid-19. Don't believe the hype. The Guardian. 2020 Apr 06. URL: https://www.theguardian.com/science/2020/apr/06/coronavirus-cure-fact-check-hydroxychloroquine-trump [accessed 2021-01-05]
9. Londoño E. Bolsonaro hails anti-malaria pill even as he fights coronavirus. The New York Times. 2020 Jul 08. URL: https://www.nytimes.com/2020/07/08/world/americas/brazil-bolsonaro-covid-coronavirus.html [accessed 2021-01-05]
10. Funke D, Flamini D. A guide to anti-misinformation actions around the world. Poynter. 2020. URL: https://www.poynter.org/ifcn/anti-misinformation-actions/ [accessed 2021-01-05]
11. Potter WJ. The state of media literacy. J Broadcast Electron Media 2010 Nov 30;54(4):675-696. [doi: 10.1080/08838151.2011.521462]
12. If Crisis or War Comes. Karlstad, Sweden: Swedish Civil Contingencies Agency; 2018 May. URL: https://tinyurl.com/y2fp39ul [accessed 2021-01-05]
13. Tips to spot false news. Facebook. 2020. URL: https://www.facebook.com/help/188118808357379 [accessed 2021-01-05]
14. Guess AM, Lerner M, Lyons B, Montgomery JM, Nyhan B, Reifler J, et al. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proc Natl Acad Sci U S A 2020 Jul 07;117(27):15536-15545 [FREE Full text] [doi: 10.1073/pnas.1920498117] [Medline: 32571950]
15. Hameleers M. Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. Inf Commun Soc 2020 May 18:1-17. [doi: 10.1080/1369118X.2020.1764603]
16. Vraga EK, Bode L, Tully M. Creating news literacy messages to enhance expert corrections of misinformation on Twitter. Communic Res 2020 Jan 30:1-23 [FREE Full text] [doi: 10.1177/0093650219898094]
17. Nyhan B, Reifler J. When corrections fail: The persistence of political misperceptions. Polit Behav 2010 Mar 30;32(2):303-330. [doi: 10.1007/s11109-010-9112-2]
18. Cook J. Countering climate science denial and communicating scientific consensus. Oxford Research Encyclopedia of Climate Science. Oxford, UK: Oxford University Press; 2016 Oct 26. URL: https://oxfordre.com/climatescience/view/10.1093/acrefore/9780190228620.001.0001/acrefore-9780190228620-e-314 [accessed 2021-01-05]
19. van der Linden S, Leiserowitz A, Maibach E. The gateway belief model: A large-scale replication. J Environ Psychol 2019 Apr;62:49-58. [doi: 10.1016/j.jenvp.2019.01.009]
20. van der Linden SL, Leiserowitz AA, Feinberg GD, Maibach EW. The scientific consensus on climate change as a gateway belief: Experimental evidence. PLoS One 2015;10(2):e0118489 [FREE Full text] [doi: 10.1371/journal.pone.0118489] [Medline: 25714347]
21. Hertwig R, Grüne-Yanoff T. Nudging and boosting: Steering or empowering good decisions. Perspect Psychol Sci 2017 Nov;12(6):973-986. [doi: 10.1177/1745691617702496] [Medline: 28792862]
22. van Stekelenburg A, Schaap G, Veling H, Buijzen M. Boosting consensus reasoning about coronavirus information (preregistration). Open Science Framework (OSF). Charlottesville, VA: Center for Open Science; 2020 Nov 04. URL: https://osf.io/u49b2 [accessed 2021-01-05]
23. van Stekelenburg A, Schaap G, Veling H, Buijzen M. Investigating and improving the accuracy of US citizens’ beliefs about the COVID-19 pandemic: A longitudinal survey study (OSF Project). Open Science Framework (OSF). Charlottesville, VA: Center for Open Science; 2020 Apr 22. URL: https://osf.io/4e9ky/ [accessed 2021-01-05]
24. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
25. ACS demographic and housing estimates. United States Census Bureau. 2019. URL: https://data.census.gov/cedsci/table?q=United%20States&g=0100000US&tid=ACSDP1Y2019.DP05 [accessed 2021-01-05]

26. Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. ArXiv Preprint posted online on March 31, 2020. [Medline: 32550244]
27. American News Pathways project: Explore the data. Pew Research Center. Washington, DC: Pew Research Center; 2020. URL: https://www.pewresearch.org/pathways-2020/covidfol/main_source_of_election_news/us_adults/ [accessed 2021-01-05]
28. Kahan DM. Ideology, motivated reasoning, and cognitive reflection. Judgm Decis Mak 2013 Jul;8(4):407-424 [FREE Full text]
29. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
30. R Core Team. The R Project for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.r-project.org/ [accessed 2021-01-05]
31. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. J Stat Softw 2017 Dec 06;82(13):1-26. [doi: 10.18637/jss.v082.i13]
32. van Holm EJ, Monaghan J, Shahar DC, Messina J, Surprenant C. The impact of political ideology on concern and behavior during COVID-19. SSRN 2020 Apr 13:1-27. [doi: 10.2139/ssrn.3573224]
33. Halekoh U, Højsgaard S. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models: The R package pbkrtest. J Stat Softw 2014;59(9):1-32. [doi: 10.18637/jss.v059.i09]
34. Morey RD. Confidence intervals from normalized data: A correction to Cousineau (2005). Quant Method Psychol 2008 Sep 01;4(2):61-64. [doi: 10.20982/tqmp.04.2.p061]
35. Hope RM. Package Rmisc: Ryan miscellaneous. The Comprehensive R Archive Network. 2016 Aug 29. URL: https://cran.r-project.org/web/packages/Rmisc/Rmisc.pdf [accessed 2021-01-05]
36. Gallagher F, Bell B. Coronavirus misinformation is widespread, according to new report that calls it an 'infodemic'. ACB News. 2020 Apr 23. URL: https://abcnews.go.com/US/coronavirus-misinformation-widespread-report-calls-infodemic/story?id=70249400 [accessed 2021-01-05]
37. O'Sullivan D. How Covid-19 misinformation is still going viral. CNN Business. 2020 May 09. URL: https://edition.cnn.com/2020/05/08/tech/covid-viral-misinformation/index.html [accessed 2021-01-05]
38. Mian A, Khan S. Coronavirus: The spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
39. Van Bavel JJ, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
40. NHS takes action against coronavirus fake news online. NHS. 2020 Mar 10. URL: https://www.england.nhs.uk/2020/03/nhs-takes-action-against-coronavirus-fake-news-online/ [accessed 2021-01-05]
41. Pennycook G, McPhetres J, Bago B, Rand D. Attitudes about COVID-19 in Canada, the UK, and the USA: A novel test of political polarization and motivated reasoning. PsyArXiv Preprints Preprint posted online on April 14, 2020. [doi: 10.31234/osf.io/zhjkp]
42. Sutton RM, Douglas KM. Agreeing to disagree: Reports of the popularity of Covid-19 conspiracy theories are greatly exaggerated. Psychol Med 2020 Jul 20:1-3. [doi: 10.1017/S0033291720002780] [Medline: 32684197]
43. Ballew M, Bergquist P, Goldberg M, Gustafson A, Kotcher J, Marlon J, et al. American Public Responses to COVID-19. New Haven, CT: Yale University and George Mason University; 2020 Apr. URL: https://climatecommunication.yale.edu/wp-content/uploads/2020/04/american-public-responses-covid19-april-2020b.pdf [accessed 2021-01-05]
44. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci 2020 Oct;7(10):201199 [FREE Full text] [doi: 10.1098/rsos.201199] [Medline: 33204475]
45. Fletcher A, Cornia A, Graves L, Kleis Nielsen R. Measuring the reach of "fake news" and online disinformation in Europe. Reuters Institute for the Study of Journalism. Oxford, UK: University of Oxford; 2018. URL: https://reutersinstitute.politics.ox.ac.uk/our-research/measuring-reach-fake-news-and-online-disinformation-europe [accessed 2021-01-05]
46. Guess AM, Nyhan B, Reifler J. Exposure to untrustworthy websites in the 2016 US election. Nat Hum Behav 2020 May;4(5):472-480 [FREE Full text] [doi: 10.1038/s41562-020-0833-x] [Medline: 32123342]
47. Allen J, Howland B, Mobius M, Rothschild D, Watts DJ. Evaluating the fake news problem at the scale of the information ecosystem. Sci Adv 2020 Apr;6(14):eaay3539 [FREE Full text] [doi: 10.1126/sciadv.aay3539] [Medline: 32284969]
48. Oreskes N, Conway EM. Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming. New York, NY: Bloomsbury Press; 2011.
49. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 2020 Jul;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]
50. Westlund O, Ghersetti M. Modelling news media use. Journal Stud 2014 Jan 07;16(2):133-151. [doi: 10.1080/1461670x.2013.868139]

51. Coronavirus Research. April 2020. Series 4: Media Consumption and Sport. London, UK: GlobalWebIndex; 2020 Apr. URL: https://www.globalwebindex.com/hubfs/1. Coronavirus Research PDFs/GWI coronavirus findings April 2020 - Media Consumption (Release 4).pdf [accessed 2021-01-05]
52. Funk C, Tyson A. Partisan differences over the pandemic response are growing. Pew Research Center. Washington, DC: Pew Research Center; 2020 Jun 03. URL: https://www.pewresearch.org/science/2020/06/03/partisan-differences-over-the-pandemic-response-are-growing/ [accessed 2021-01-05]
53. Leeper TJ, Slothuus R. Political parties, motivated reasoning, and public opinion formation. Polit Psychol 2014 Jan 22;35:129-156. [doi: 10.1111/pops.12164]
54. Slater MD. Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. Media Psychol 2015 Jul 01;18(3):370-395 [FREE Full text] [doi: 10.1080/15213269.2014.897236] [Medline: 26366124]
55. Goodier M. Infodemic investigation: Facebook is biggest source of false claims about coronavirus. Press Gazette. 2020 Jul 06. URL: https://www.pressgazette.co.uk/infodemic-investigation-facebook-is-biggest-source-of-false-claims-about-coronavirus/ [accessed 2021-01-05]
56. De Domenico M, Castaldo N. COVID19 infodemics observatory. Open Science Framework (OSF). Charlottesville, VA: Center for Open Science; 2020 Mar 17. URL: https://osf.io/n6upx/ [accessed 2021-01-05]
57. Merkley E. Are experts (news)worthy? Balance, conflict, and mass media coverage of expert consensus. Polit Commun 2020 Feb 17;37(4):530-549. [doi: 10.1080/10584609.2020.1713269]
58. Shorman J, Chambers F. Kansas official says man drank cleaner after Trump floated dangerous disinfectant remedy. The Wichita Eagle. 2020 Apr 27. URL: https://www.kansas.com/news/politics-government/article242326966.html [accessed 2021-01-05]
59. Burki T. The online anti-vaccine movement in the age of COVID-19. Lancet Digit Health 2020 Oct;2(10):e504-e505. [doi: 10.1016/s2589-7500(20)30227-2]

Abbreviations

OSF: Open Science Framework
T0: baseline
T1: first follow-up wave
T2: second follow-up wave
T3: final wave
WHO: World Health Organization

Edited by G Eysenbach; submitted 02.09.20; peer-reviewed by E Vraga, J Feliciano, H Bhatt, A Azzam, Z Ren, N Mohammad Gholi Mezerji, C Miranda; comments to author 26.09.20; revised version received 02.11.20; accepted 07.12.20; published 12.01.21.

Please cite as:
van Stekelenburg A, Schaap G, Veling H, Buijzen M
Investigating and Improving the Accuracy of US Citizens' Beliefs About the COVID-19 Pandemic: Longitudinal Survey Study
J Med Internet Res 2021;23(1):e24069
URL: http://www.jmir.org/2021/1/e24069/
doi:10.2196/24069
PMID:33351776

Original Paper

Digital Health Literacy and Web-Based Information-Seeking Behaviors of University Students in Germany During the COVID-19 Pandemic: Cross-sectional Survey Study

Kevin Dadaczynski[1,2*], MSc, PhD; Orkan Okan[3*], PhD; Melanie Messer[4], PhD; Angela Y M Leung[5], BN, MHA, PhD; Rafaela Rosário[6], MSc, PhD; Emily Darlington[7], MPH, MSc, PhD; Katharina Rathmann[1], PhD

[1]Department of Nursing and Health Sciences, Fulda University of Applied Sciences, Fulda, Germany

[2]Centre for Applied Health Science, Leuphana University Lueneburg, Lueneburg, Germany

[3]Interdisciplinary Centre for Health Literacy Research, Faculty of Educational Science, Bielefeld University, Bielefeld, Germany

[4]Section "External Lecturer", APOLLON University of Applied Sciences, Bremen, Germany

[5]Centre for Gerontological Nursing, World Health Organization Collaborating Centre for Community Health Services, School of Nursing, Hong Kong Polytechnic University, Hong Kong SAR, Hong Kong, China

[6]School of Nursing, University of Minho, Braga, Portugal

[7]Health Systemic Process EA 4129 Research Unit (P2S), University of Lyon, University Claude Bernard of Lyon 1, Lyon, France

[*]these authors contributed equally

Corresponding Author:
Kevin Dadaczynski, MSc, PhD
Department of Nursing and Health Sciences
Fulda University of Applied Sciences
Leipziger Sraße 123
Fulda, 36037
Germany
Phone: 49 661 9640 6073
Email: kevin.dadaczynski@pg.hs-fulda.de

Abstract

Background: Digital communication technologies are playing an important role in the health communication strategies of governments and public health authorities during the COVID-19 pandemic. The internet and social media have become important sources of health-related information on COVID-19 and on protective behaviors. In addition, the COVID-19 infodemic is spreading faster than the coronavirus itself, which interferes with governmental health-related communication efforts. This jeopardizes national public health containment strategies. Therefore, digital health literacy is a key competence to navigate web-based COVID-19–related information and service environments.

Objective: This study aimed to investigate university students' digital health literacy and web-based information-seeking behaviors during the early stages of the COVID-19 pandemic in Germany.

Methods: A cross-sectional study among 14,916 university students aged ≥18 years from 130 universities across all 16 federal states of Germany was conducted using a web-based survey. Along with sociodemographic characteristics (sex, age, subjective social status), the measures included five subscales from the Digital Health Literacy Instrument (DHLI), which was adapted to the specific context of the COVID-19 pandemic. Web-based information-seeking behavior was investigated by examining the web-based sources used by university students and the topics that the students searched for in connection with COVID-19. Data were analyzed using univariate and bivariate analyses.

Results: Across digital health literacy dimensions, the greatest difficulties could be found for assessing the reliability of health-related information (5964/14,103, 42.3%) and the ability to determine whether the information was written with a commercial interest (5489/14,097, 38.9%). Moreover, the respondents indicated that they most frequently have problems finding the information they are looking for (4282/14,098, 30.4%). When stratified according to sociodemographic characteristics, significant differences were found, with female university students reporting a lower DHLI for the dimensions of "information searching" and "evaluating reliability." Search engines, news portals, and websites of public bodies were most often used by the respondents as sources to search for information on COVID-19 and related issues. Female students were found to use social media and health portals more frequently, while male students used Wikipedia and other web-based encyclopedias as well as YouTube more often. The use of social media was associated with a low ability to critically evaluate information, while the opposite was observed for the use of public websites.

Conclusions: Although digital health literacy is well developed in university students, a significant proportion of students still face difficulties with certain abilities to evaluate information. There is a need to strengthen the digital health literacy capacities of university students using tailored interventions. Improving the quality of health-related information on the internet is also key.

(J Med Internet Res 2021;23(1):e24097) doi:10.2196/24097

KEYWORDS

digital health; literacy; infodemic; health information; behaviour; coronavirus; COVID-19; university student; student; infodemiology

Introduction

Shortly after the outbreak of SARS-CoV-2 and the associated disease, COVID-19, were first reported [1], it was declared a pandemic [2,3] by the World Health Organization. When the first case of COVID-19 was reported in Germany on January 27, 2020, the government responded immediately by launching an unprecedented nationwide emergency response plan that focused on four pillars: prevention, detection, containment, and treatment [4]. In addition to the National Pandemic Plan [5] and to health care and medical interventions [4], the government endorsed a public health communication strategy that was supported by all health agencies and public health bodies on national and local levels [6-9]. This communication strategy involved public broadcasting agencies, which launched web-based media campaigns, including daily nationwide podcasts. The underlying objective of this approach was to provide citizens with the necessary information on COVID-19 and how it affected people's health [4,10-13]. Within a short period of time, a massive amount of web-based health-related information on COVID-19 became available on issues such as protective behaviors, preventive measures, treatment options, dashboard statistics, the latest scientific insights, and various safety recommendations [14-16]. It has now become clear that this pandemic has been accompanied by an "infodemic"—an overabundance of valid and invalid health information on COVID-19 [17,18]. By means of digital communication technologies, especially the internet and social media, the COVID-19 infodemic is spreading faster than the coronavirus itself, which interferes with governmental health communication efforts and jeopardizes national public health containment strategies.

Altogether, this situation creates a complex information environment that requires people to be able to access, navigate, understand, use, and critically evaluate information and services in ways that support healthy and protective behaviors in the time of the COVID-19 pandemic. Therefore, health literacy, which is the ability to find, understand, and evaluate health information and apply it in daily decision-making and health behavior [19], is of utmost importance during the current pandemic [14]. Digital health literacy applies this understanding of health literacy to digital contexts and environments [20], and it has become a core competence and necessity for navigating web-based information and health service environments within the realm of the COVID-19 pandemic and the associated infodemic [21]. However, in Germany, it has been shown that more than half of the population has limited health literacy; therefore, people report difficulties in dealing with health-related information [22]. A recent study conducted in Germany on health literacy in relation to information regarding COVID-19 resulted in similar findings [23]. People particularly have difficulty assessing the trustworthiness of media information on COVID-19 and its associated health problems. In addition, people with limited health literacy are more likely to be confused due to the massive amounts of information available in the media and on the internet [23]. Information is a carrier of important health knowledge to contain the virus and empower citizens to demonstrate health literacy [16,24]; the pandemic has placed increased demand on the general population to find information relevant to them and critically reflect on this information, as well as to transfer information into their everyday life and practices.

This issue is particularly critical for university students, who consist of a significant proportion of young adults in Germany. University students comprise the population that primarily uses digital technologies and web-based health information [25,26]. Although it can be noted that students have not been the primary focus of research since the beginning of the pandemic, a recent study with over 5400 medical students from Vietnam revealed that higher levels of health literacy were associated with less fear of COVID-19 [27]. Therefore, health literacy is a critical intervention target, especially since fear is one of the toxic outcomes that result from an infodemic [17,28]. The aim of this study is to investigate the digital health literacy and web-based information-seeking behaviors among university students in Germany during the first wave of the COVID-19 pandemic, particularly during university closures. This study is informed by the conceptual model of health literacy as presented by Sørensen and colleagues [19] and the model of digital health literacy as proposed by van der Vaart and Drossaert [20]. The assumptions in both models are that personal and environmental determinants influence an individual's capacity regarding various dimensions of personal information management, which include informing health decisions and behaviors that are beneficial for health. Our study focuses on personal and environmental determinants, personal information management, and behavioral aspects. In this context, the following research questions were addressed:

- What are the levels of COVID-19–related digital health literacy in German university students stratified by social, economic, and geographical indicators?

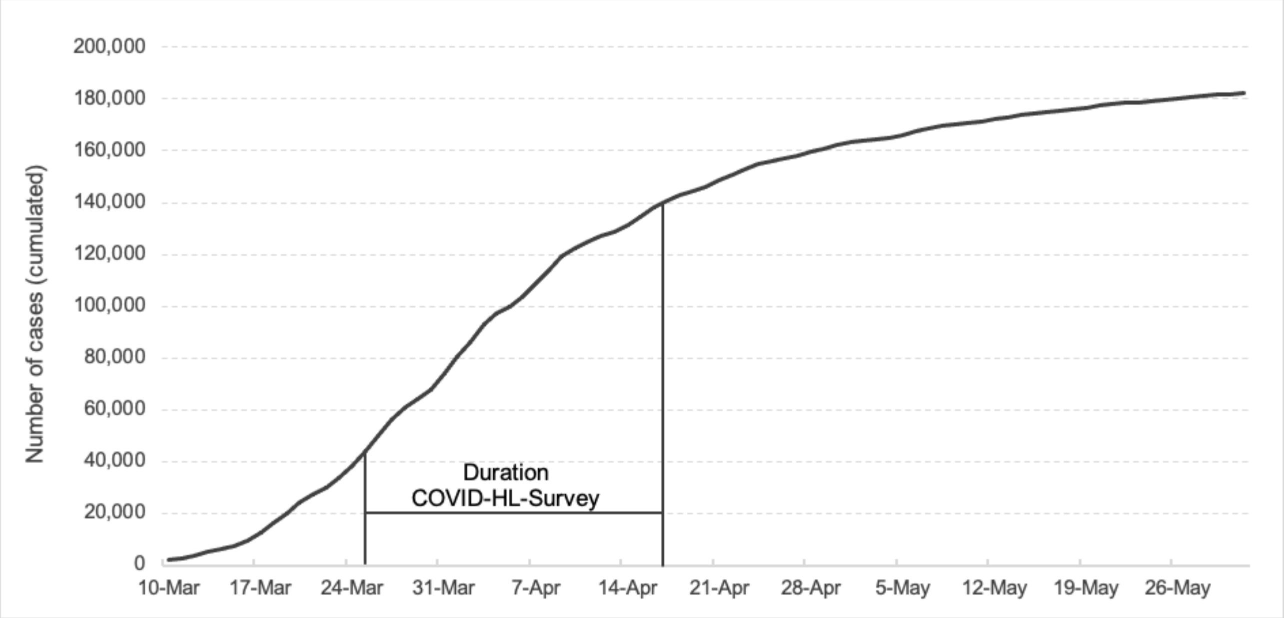
- Which sources of web-based information are used and which topics are searched for in the context of COVID-19 by German university students?
- Can differences be identified between students with regard to health literacy, the sources used for information searching, and the topics addressed in relation to COVID-19?

Methods

Study Design and Participants

A national cross-sectional web-based survey was conducted including a nonrandomized sample (convenience sample) of German university students. To address as many university students as possible, all private and state universities (ie, 392 universities containing 2.9 million students [29]) were invited to participate in the study by email. A reminder was sent two weeks after the survey started. The presidencies of all the universities and the deaneries of all faculties were contacted and asked to forward an invitation letter to their students. University students enrolled at a private or state university were eligible to participate in this study. To increase the homogeneity of the sample, respondents were initially asked to indicate their current status. Those who indicated that they were not currently enrolled as students at a German university were excluded from the data set (n=245). The duration of the study was 3 weeks, and it took place from March 25 to April 17, 2020. Within the 3 weeks during which the survey was implemented, the number of confirmed COVID-19 cases in Germany increased from 44,175 to 141,016 (Figure 1 [30]). The survey was administered electronically using the Enterprise Feedback Suite survey tool (Questback) [31]. Participation was voluntary, and anonymity was ensured. Upon entering the web-based survey site, participants were presented with information regarding the background and the aims of the study. After checking a consent box at the bottom of the page, participants were directed to the questionnaire. Our study was approved by the Bielefeld University ethics committee (No. EUB 2020-053).

Figure 1. Confirmed cumulative cases of COVID-19 in Germany from March to May 2020 (source: RKI COVID-19 Dashboard [30]).



Measures

Sociodemographic information included sex (male, female, diverse), age, study course (bachelor's degree, master's degree, other), and subjective social status (SSS). Age was measured in absolute numbers, and based on an analysis of the distribution, four categories were created (≤20 years, 21-23 years, 24-26 years, and ≥27 years). Social status was assessed using the German version of the MacArthur Scale, which includes a ladder with 10 steps [32]. Respondents were asked to position themselves at the step that best reflected their status in the social hierarchy, with higher values indicating a higher social status. According to previous studies, respondents were categorized into three groups: low SSS (1-4), medium SSS (5-7), and high SSS (8-10) [33].

Digital health literacy was evaluated using five of the seven subscales from the validated Digital Health Literacy Instrument (DHLI) [20], each including three items to be answered on a 4-point scale (eg, 1, very difficult; 4, very easy). The DHLI was adapted to the context of the COVID-19 pandemic (eg, "When you search the Internet for information on the coronavirus or related topics, how easy or difficult is it for you to…"). The five subscales include (1) searching the web for information on COVID-19, (2) adding self-generated content on COVID-19, (3) evaluating the reliability of COVID-19–related information, (4) determining personal relevance of COVID-19–related information, and (5) protecting privacy on the internet. The internal consistency (Cronbach α) of the first four subscales was acceptable to good ($.70<\alpha<.83$). Due to low reliability ($\alpha=.46$), scaling was omitted for the protecting privacy subscale.

The section about web-based information-seeking behaviors focused on the sources that were used to receive web-based health information about COVID-19 and related topics. The respondents were presented with a 10-item list of different web-based sources (eg, search engines, websites of public health bodies, government agencies, and social media providers), in which the frequency of their use could be rated on a 5-point scale (0, don't know; 4, often) [34]. Students were also asked to indicate the specific topics they searched for in the context of COVID-19. The assessment was based on a self-developed list of 9 topics (eg, current spread of COVID-19, symptoms of COVID-19, measures to protect against infection, dealing with psychological stress caused by the COVID-19 pandemic), using yes or no answers.

An overview of all items and scales used for this paper can be found in Multimedia Appendix 1. The entire questionnaire is available on request from the first authors.

Statistical Analysis

To control for the selection bias caused by a convenience sampling procedure, we used weighting to adjust the sample distribution to the characteristics of the general population of German university students. Based on the data provided by the Federal Statistical Office via the GENESIS database [35], the data could be weighted for gender and desired study degree. In the first step, all data on digital health literacy and information-seeking behavior were analyzed descriptively. Subsequently, bivariate analyses were conducted by cross-tabulating the two levels of digital health literacy (limited vs sufficient) with sociodemographic characteristics using chi-square tests. For this purpose, all DHLI subscales (except “protecting privacy”) were dichotomized using median splits. Due to the low internal consistency for the dimension “protecting privacy” and the fact that two subscales from the original DHLI instrument were not used, we also refrained from calculating an overall mean value, as done by Van der Vaart and Drossaert [20]. For all analyses, *P* values <.05 were considered statistically significant. However, due to the large sample size, the strength of the association was determined using the Cramer index (Cramer V). The Cramer V is a normalized version of the chi square statistic test for nominal scaled variables. According to Cohen [36], the strength of each association was interpreted as an effect size measure using the following conventions: ≥0.1 (small), ≥0.3 (medium), ≥0.5 (large). In further analyses, chi-square tests were also performed for the levels of digital health literacy and the topics searched for with regard to COVID-19. Finally, to analyze differences between the levels of digital health literacy and the sources used to search for COVID-19–related information, *t* tests for independent samples were conducted. Cohen *d* was used as an effect size measure by applying the following conventions: ≥0.2 (small), ≥0.5 (medium), ≥0.8 (large) [36].

Results

After further plausibility checks and adjustment for incorrect data, the data set contained complete questionnaires from 14,916 participants aged between 18 and 72 years (mean age: 24.3). Students from 130 universities and all 16 federal states participated (see Table 1). In terms of geographical coverage (see Figure 2), most respondents were from the west of Germany (6355/14,833, 42.8%), followed by students from the south (3694/14,833, 24.9), and almost equally from the north (2307/14,833, 15.6%) and the east (2476/14,833, 16.7%).

Table 1. Characteristics of the study participants (N=14,916; values are weighted).

Characteristic	Value, n (%)		
	Total	Male (n=7687, 51.5%)	Female (n=7229, 48.5%)
Age (years; n=14,897)			
≤20	2640 (17.7)	1342 (17.5)	1298 (18.0)
21-23	5495 (36.9)	2586 (33.7)	2909 (40.3)
24-26	3567 (23.9)	1923 (25.0)	1643 (22.8)
≥27	3195 (21.4)	1827 (23.8)	1369 (19.0)
Study course (n=14,916)			
Bachelor’s degree	10,351 (69.4)	5463 (71.1)	4887 (67.6)
Master’s degree	2796 (18.7)	1460 (19.0)	1337 (18.5)
Other (eg, PhD)	1769 (11.9)	764 (9.9)	1005 (13.9)
Subjective social status (n=14,913)			
Low	2575 (17.3)	1408 (18.3)	1168 (16.2)
Middle	10,090 (67.7)	4974 (64.7)	5116 (70.8)
High	2247 (15.1)	1303 (17.0)	945 (13.1)

Figure 2. Geographical distribution of the study sample (N=14,916). BB: Brandenburg; BE: Berlin; BW: Baden-Wuerttemberg, BY: Bavaria; HB: Bremen; HE: Hesse; HH: Hamburg; MV: Mecklenburg-Western Pomerania; NI: Lower Saxony; NW: North Rhine-Westphalia; RP: Rhineland-Palatinate; SH: Schleswig-Holstein; SL=Saarland; SN: Saxony; ST: Saxony Anhalt; TH: Thuringia.

In comparison with the whole population of German university students via the GENESIS database, some deviations could be observed. While students from Baden-Wuerttemberg (12.7% vs 4.5%) and North Rhine-Westphalia (27.5% vs 12.4%) were underrepresented in our study, our sample includes significantly more students from Bavaria (14.0% vs 20.4%) and Hesse (9.4% vs 21.7%). The gender distribution was almost balanced, with 51.5% male university students (7687/14,916) and 48.5% female students (7229/14,913). Regarding SSS, more than two-thirds of respondents reported a middle SSS (10,090/14,913, 67.7%), while 17.3% (2575/14,916) reported a low SSS and 15.1% (2247/14,913) reported a high SSS (mean SSS 6.0, SD 1.54).

Figures 3 to 7 show the different dimensions of digital health literacy and the percentages of student scoring. Within the "information search" subscale, university students indicated that they most frequently had problems finding the information they were looking for (4282/14,098, 30.4%), while the use of suitable words and search queries caused less difficulty (1644/14,101, 11.7%). Regarding the dimension of "adding self-generated content," respondents reported the most difficulties in expressing their own opinion, in expressing thoughts or feelings in writing (3975/13,754, 28.9%), and in writing a message in a way that is understandable for others (4661/13,752, 33.9%). Across all dimensions, the greatest difficulties could be found in assessing the reliability of health-related information (5964/14,103, 42.3%) and the ability to determine whether the information was written with commercial interest (5489/14,097, 38.9%). The use of the found information for one's own health-related decisions (eg, regarding protective measures, 2443/14,079, 14.4%) and the application of this information in daily life caused difficulties for approximately one-fifth of the respondents (2812/14,067, 20.0%). Finally, some heterogeneity could also be found in the items relating to the dimension of "protecting privacy." Although approximately 35% of the respondents experienced difficulties to judge who could read messages posted on the web (4768/13,589, 35.1%), only 6.7% stated that they sometimes or often shared private information on the web (914/13,715).

Figure 3. Responses to questions in the Digital Health Literacy Instrument subscale “information search” (n=14,098 to n=14,110), %.

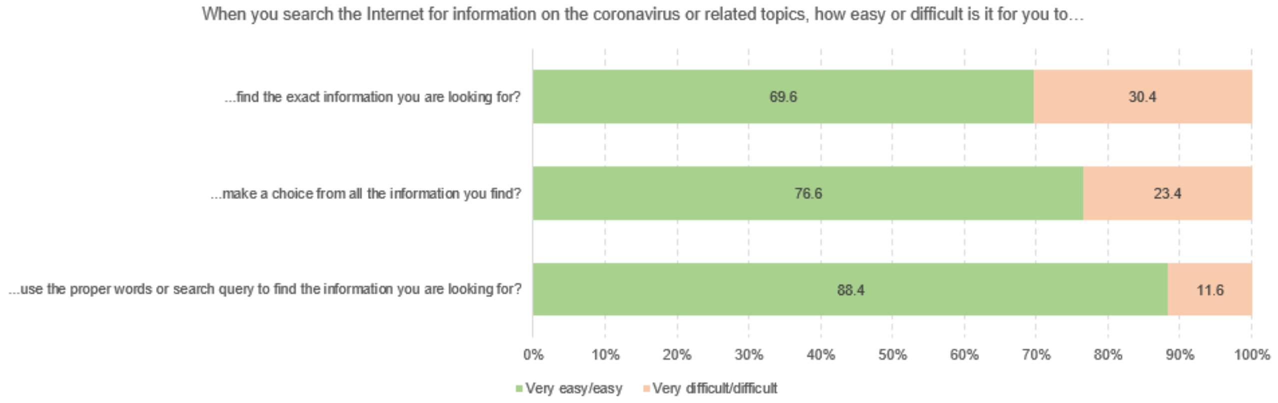


Figure 4. Responses to questions in the Digital Health Literacy Instrument subscale “adding self-generated content” (n=13,721 to n=13,754), %.

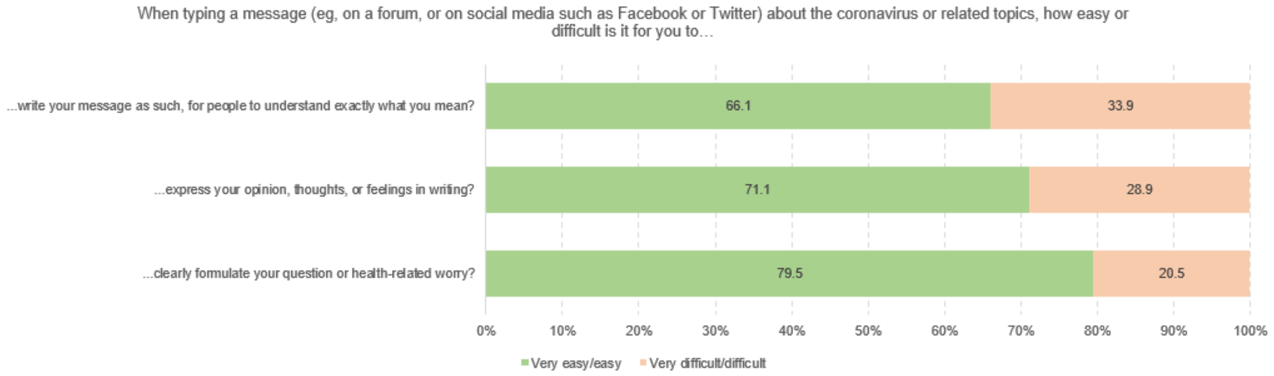


Figure 5. Responses to questions in the Digital Health Literacy Instrument subscale “evaluating reliability” (n=14,081 to n=14,103), %.

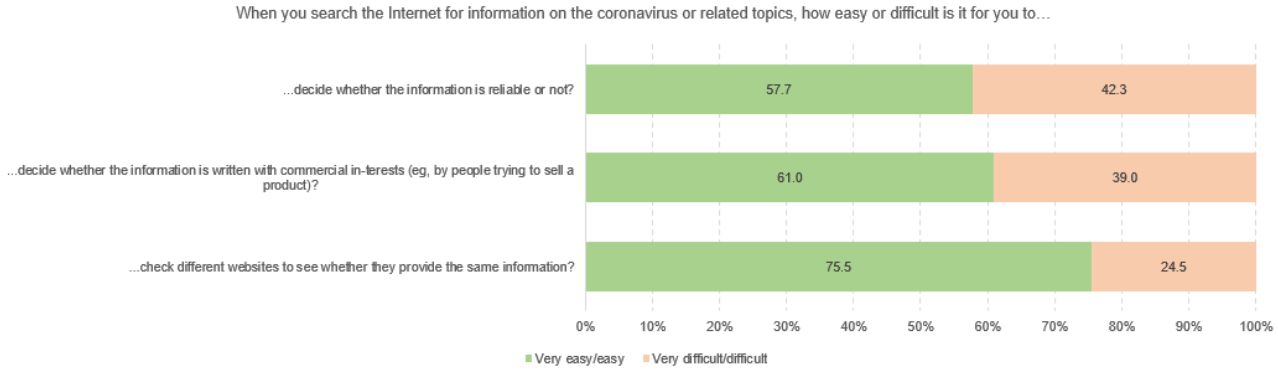


Figure 6. Responses to questions in the Digital Health Literacy Instrument subscale “determining relevance” (n=14,076 to n=14,092), %.

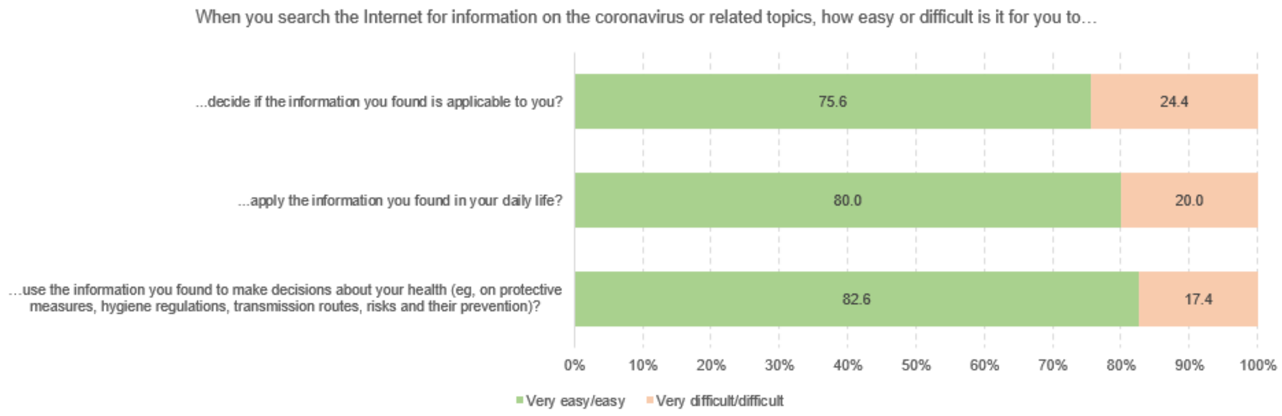


Figure 7. Responses to questions in the Digital Health Literacy Instrument subscale “protecting privacy” (n=13,589 to n=13,715), %.

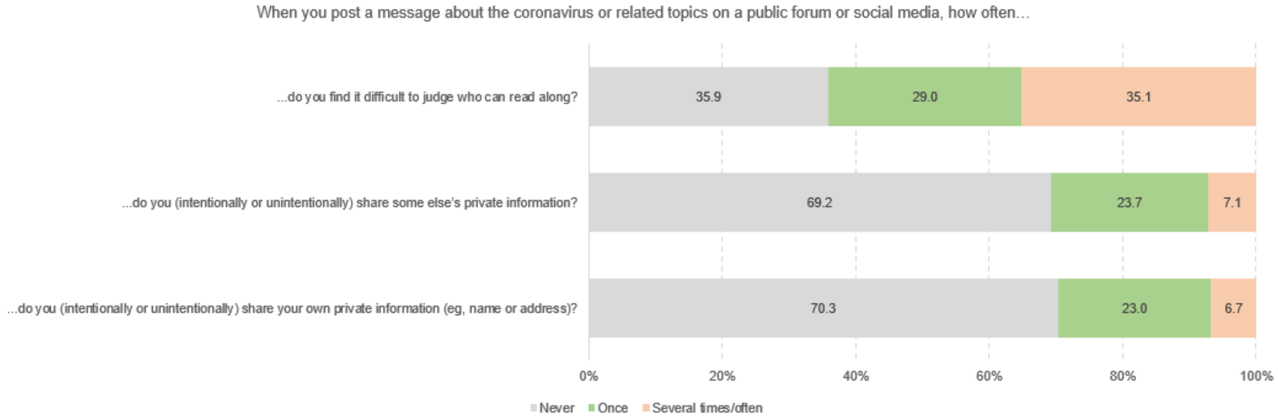


Table 2 and Table 3 show the digital health literacy levels of the respondents, stratified by sociodemographic and geographic characteristics. Concerning gender, significant differences were found, with female university students showing lower digital health literacy across all subscales. However, taking the strength of the association (*V*) into account, small effect sizes could be identified only for the dimensions “information searching” (male: 2087/7219, 28.9%, female: 2711/6865, 39.5%, χ^2_1=175.37, *P*<.001, *V*=0.11) and “evaluating reliability” (male: 2156/5994, 36.0%, female: 2660/5630, 47.2%, χ^2_1=152.16, *P*<.001, *V*=0.11). All other differences were below the threshold for small effects and were hence considered trivial. When considering differentiation by age group, in all subscales, a slight tendency of increasing level of digital health literacy with increasing age was observed. However, these significant differences proved to be trivial when calculating effect sizes. The same was observed for study course, SSS, and geographical distribution. Slight differences between the groups were observed; however, the differences remained below the threshold for small effects.

Table 2. Digital health literacy levels of university students for the subscales of "information search" and "adding self-generated content" according to their sociodemographic and geographic characteristics.

Characteristic	Information search					Adding self-generated content				
	Limited, n (%)	Sufficient, n (%)	χ^2 (*df*)	*P*	V	Limited, n (%)	Sufficient, n (%)	χ^2 (*df*)	*P*	V
Gender			175.37 (1)	<.001	0.11			29.77 (1)	<.001	0.05
Male	2087 (28.9)	5132 (71.1)				2687 (38.7)	4255 (61.3)			
Female	2711 (39.5)	4154 (60.5)				2913 (43.3)	3815 (56.7)			
Age (years)			17.77 (3)	<.001	0.04			78.38 (3)	<.001	0.08
≤20	868 (35.3)	1593 (64.7)				1091 (45.7)	1295 (54.3)			
21-23	1841 (35.5)	3340 (64.5)				2174 (43.0)	2883 (57.0)			
24-26	1139 (33.6)	2250 (66.4)				1307 (39.8)	1975 (60.2)			
≥27	948 (31.2)	2088 (68.8)				1020 (34.8)	1909 (65.2)			
Study course			26.17 (2)	<.001	0.04			98.48 (2)	<.001	0.08
Bachelor's degree	3431 (35.4)	6253 (64.6)				5339 (56.6)	4087 (43.4)			
Master's degree	848 (31.5)	1845 (68.5)				1577 (61.2)	1001 (38.8)			
Other (eg, PhD)	519 (30.4)	1188 (69.6)				1153 (69.2)	513 (30.8)			
Subjective social status			11.84 (2)	.003	0.03			29.64 (2)	<.001	0.05
Low	831 (34.5)	1580 (65.5)				1010 (43.1)	1335 (56.9)			
Middle	3310 (34.7)	6233 (65.3)				3853 (41.6)	5409 (58.4)			
High	655 (30.8)	1471 (69.2)				735 (35.7)	1325 (64.3)			
Geographic location			39.15 (3)	<.001	0.05			39.28 (3)	<.001	0.05
North	686 (31.6)	1484 (68.4)				857 (40.6)	1255 (59.4)			
East	786 (33.5)	1561 (66.5)				582 (37.7)	1406 (62.3)			
West	1962 (32.7)	4036 (67.3)				2325 (39.9)	3508 (60.1)			
South	1338 (38.3)	2154 (61.7)				1539 (45.3)	1858 (54.7)			
Total	4798 (34.1)	9286 (65.9)				5600 (41.0)	8069 (59.0)			

Table 3. Digital health literacy levels of university students for the subscales of "evaluating reliability" and "determining relevance" according to their sociodemographic and geographic characteristics.

Characteristic	Evaluating reliability					Determining relevance				
	Limited, n (%)	Sufficient, n (%)	χ^2 (*df*)	*P*	V	Limited, n (%)	Sufficient, n (%)	χ^2 (*df*)	*P*	V
Gender			152.16 (1)	<.001	0.11			38.10 (1)	<.001	0.05
Male	2156 (36.0)	3838 (64.0)				2387 (33.2)	4799 (66.8)			
Female	2660 (47.2)	2970 (52.8)				2617 (38.2)	4232 (61.8)			
Age (years)			35.04 (3)	<.001	0.05			9.78 (3)	.02	0.03
≤20	864 (42.5)	1169 (57.5)				874 (35.7)	1571 (64.3)			
21-23	1825 (43.5)	2373 (56.5)				1859 (36.0)	3308 (64.0)			
24-26	1179 (42.3)	1611 (57.7)				1252 (37.1)	2124 (62.9)			
≥27	945 (36.5)	1645 (63.5)				1013 (33.4)	2017 (66.6)			
Study course			23.75 (2)	<.001	0.05			9.94 (2)	.007	0.03
Bachelor's degree	3434 (42.9)	4565 (57.1)				3513 (36.4)	6137 (63.6)			
Master's degree	846 (38.1)	1373 (61.9)				935 (34.8)	1749 (65.2)			
Other (eg, PhD)	536 (38.1)	870 (61.9)				555 (32.6)	1146 (67.4)			
Subjective social status			18.69 (2)	<.001	0.04			30.11 (2)	<.001	0.05
Low	890 (44.0)	1133 (56.0)				928 (38.6)	1475 (61.4)			
Middle	3258 (41.7)	4553 (58.3)				3421 (36.0)	6091 (64.0)			
High	665 (37.2)	1121 (62.8)				655 (30.9)	1462 (69.1)			
Geographic location			41.08 (3)	<.001	0.06			10.43 (3)	.015	0.03
North	674 (38.0)	1101 (62.0)				731 (33.8)	1433 (66.2)			
East	763 (39.1)	1187 (60.9)				855 (36.6)	1482 (63.4)			
West	2017 (40.8)	2921 (59.2)				2086 (34.9)	3897 (65.1)			
South	1339 (46.2)	1558 (53.8)				1300 (37.4)	2176 (62.6)			
Total	4816 (41.4)	6808 (58.6)				5004 (35.7)	9032 (64.3)			

Search engines, news portals, and websites of public bodies were most often used by the respondents as sources to search for and find information on COVID-19 and related issues (see Figure 8 and Figure 9). These sources were followed by social media platforms such as Facebook, Instagram, and Twitter, or video portals such as YouTube, with 37.6% of respondents (5302/14,092) stating that they used these media sometimes or frequently. In contrast, health-related blogs or web-based guides were used much less frequently. When stratified by sociodemographic characteristics, relevant differences could only be observed for gender. Female students were found to use social media ($t_{13,921}$=–19.09, *P*<.001, *d*=–0.32) and health portals ($t_{13,463}$=–14.42, *P*<.001, *d*=–0.24) more frequently than male students. In contrast, Wikipedia and other web-based encyclopedias ($t_{14,051}$=19.19, *P*<.001, *d*=0.32), as well as YouTube ($t_{14,054}$=18.13, *P*<.001, *d*=0.30), were more often used by male students. Regarding the topics, respondents stated that they most frequently searched for information on the current spread of SARS-CoV-2 (12,648/14,114, 89.6%) and associated restrictions (12,126/14,114, 85.9%), recommendations and assessments regarding the situation (10,975/14,114, 77.8%), and the symptoms of COVID-19 (10,089/14,114, 71.5%). Although significantly less often, one-fifth of the university students stated that they looked for information on how to cope with psychological stress caused by the COVID-19 situation (2921/14,114, 20.7%). When differentiated by sociodemographic variables, gender differences could be found, as male students searched significantly more often for information on economic and social consequences of the COVID-19 pandemic (male: 4943/7237, 68.3%, female: 3817/6878, 55.5%, χ^2_1=245.62, *P*<.001, *V*=.13).

Figure 8. Internet search queries related to COVID-19 (n=14,111), %.

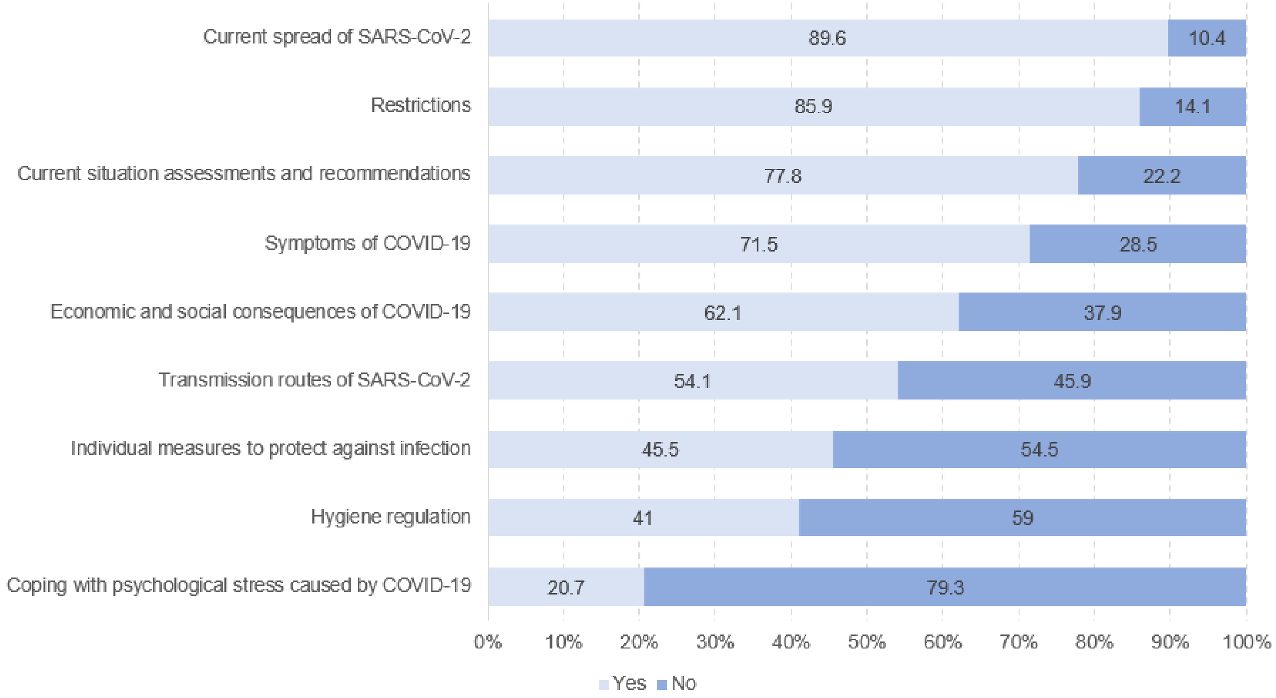
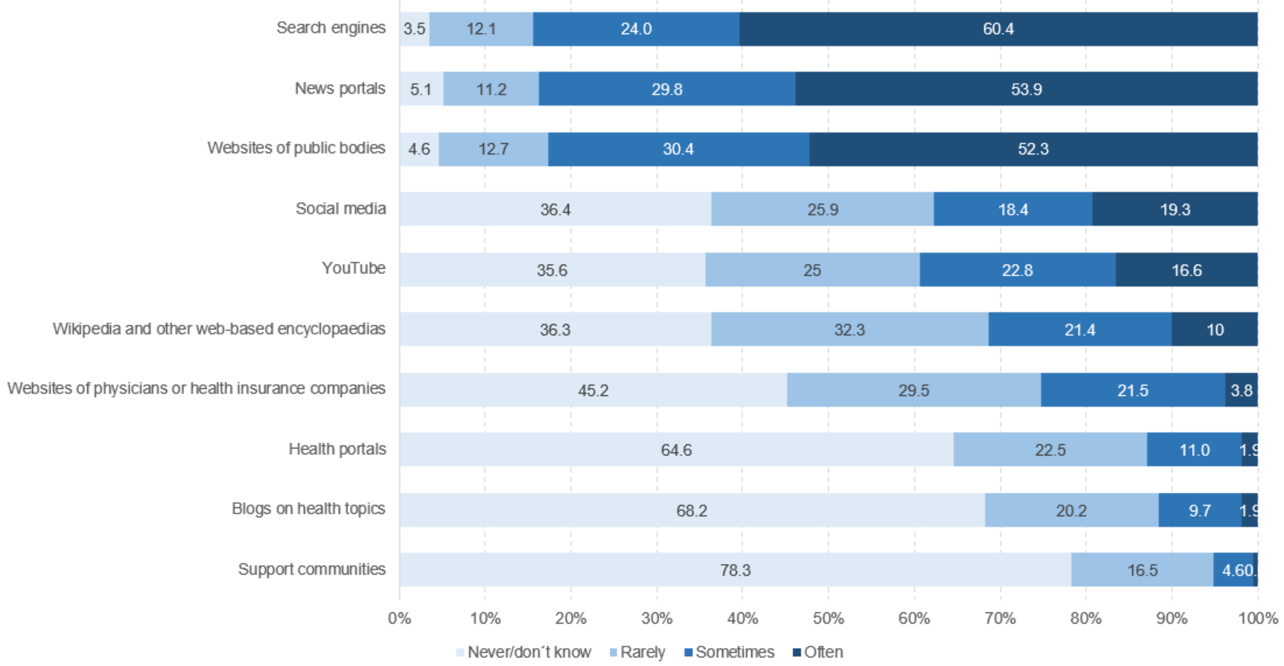


Figure 9. Frequency of use of internet sources for web-based health information seeking (n=14,012 to n=14,094), %.



Finally, digital health literacy was stratified according to web-based information-seeking behavior. No relevant differences could be found for the topics that students searched for. Regarding the sources used for the search and the handling of health-related information, significant and relevant differences emerged for the DHLI subscale "evaluating reliability." Respondents with sufficient digital health literacy in that dimension reported using the websites of public bodies (eg, Robert Koch Institute) more frequently (t_{9344}=19.44, *P*<.001, *d*=0.37). The opposite could be observed for social media (Facebook, Instagram, Twitter) ($t_{10,019}$=–14.29, *P*<.001, *d*=–0.27) and support communities (t_{9028}=–12.06, *P*<.001, *d*=–0.23), which were more frequently used by respondents who reported more difficulties in evaluating the reliability of information (see Table 4 and Table 5).

Table 4. Sources used for COVID-19 information search stratified by digital health literacy level for the "information search" and "adding self-generated content" subscales.

Item	Information search				Adding self-generated content			
	Limited, mean (SD)	Sufficient, mean (SD)	*P*	*d*	Limited, mean (SD)	Sufficient, mean (SD)	*P*	*d*
Search engines (eg, Google, Bing, Yahoo!)	3.44 (0.79)	3.40 (0.85)	.003	–0.05	3.43 (0.82)	3.42 (0.83)	.21	N/A[a]
Websites of public bodies (eg, Robert Koch Institute)	3.20 (0.88)	3.36 (0.85)	<.001	0.18	3.24 (0.87)	3.35 (0.85)	<.001	0.12
Wikipedia and other web-based encyclopedias	1.99 (0.96)	2.08 (0.99)	<.001	0.09	2.03 (0.98)	2.07 (0.98)	.005	0.05
Social media (eg, Facebook, Instagram, Twitter)	2.31 (1.14)	2.15 (1.12)	<.001	–0.14	2.29 (1.14)	2.17 (1.12)	<.001	–0.10
YouTube	2.19 (1.09)	2.21 (1.10)	.22	N/A	2.25 (1.10)	2.18 (1.09)	<.001	–0.07
Blogs on health topics	1.47 (0.74)	1.44 (0.75)	.056	N/A	1.46 (0.74)	1.45 (0.75)	.35	N/A
Support- communities	1.30 (0.59)	1.26 (0.56)	<.001	–0.08	1.30 (0.59)	1.27 (0.56)	.002	–0.05
Health portals	1.53 (0.77)	1.49 (0.76)	.004	–0.05	1.51 (0.77)	1.51 (0.77)	.80	N/A
Websites of physicians or health insurance companies	1.84 (0.89)	1.84 (0.89)	.72	N/A	1.82 (0.88)	1.86 (0.91)	.005	0.05
News portals (eg, newspapers, television)	3.30 (0.86)	3.34 (0.87)	.03	.04	3.32 (0.86)	3.32 (0.87)	.75	N/A

[a] N/A: not applicable due to lack of significance.

Table 5. Sources used for COVID-19 information search stratified by digital health literacy level for the "evaluating reliability" and "determining relevance" subscales.

Item	Evaluating reliability				Determining relevance			
	Limited, mean (SD)	Sufficient, mean (SD)	*P*	*d*	Limited, mean (SD)	Sufficient, mean (SD)	*P*	*d*
Search engines (eg, Google, Bing, Yahoo!)	3.49 (0.79)	3.36 (0.86)	<.001	–0.15	3.43 (0.82)	3.41 (0.84)	.07	N/A[a]
Websites of public bodies (eg, Robert Koch Institute)	3.13 (0.92)	3.45 (0.79)	<.001	0.37	3.22 (0.89)	3.35 (0.84)	<.001	0.16
Wikipedia and other web-based encyclopedias	2.03 (0.98)	2.08 (1.00)	.008	0.05	2.02 (0.97)	2.06 (0.99)	.01	0.04
Social media (eg, Facebook, Instagram, Twitter)	2.38 (1.15)	2.07 (1.10)	<.001	–0.27	2.28 (1.14)	2.16 (1.12)	<.001	–0.10
YouTube	2.24 (1.12)	2.17 (1.09)	.001	–0.06	2.22 (1.10)	2.19 (1.09)	.20	N/A
Blogs on health topics	1.49 (0.76)	1.42 (0.74)	<.001	–0.10	1.46 (0.74)	1.45 (0.75)	.37	N/A
Support-communities	1.36 (0.64)	1.22 (0.53)	<.001	–0.23	1.31 (0.61)	1.25 (0.55)	<.001	–0.10
Health portals	1.56 (0.80)	1.47 (0.74)	<.001	–0.12	1.52 (0.78)	1.49 (0.76)	.02	–0.04
Websites of physicians or health insurance companies	1.81 (0.88)	1.87 (0.91)	.001	0.05	1.82 (0.88)	1.85 (0.90)	.09	N/A
News portals (eg, newspapers, television)	3.30 (0.87)	3.33 (0.87)	.04	0.04	3.33 (0.84)	3.32 (0.88)	.32	N/A

[a] N/A: not applicable due to lack of significance.

Discussion

To our knowledge, this study is the first to investigate the digital health literacy and information-seeking behaviors in university students during the COVID-19 outbreak in Germany. Nationwide and overall, university students show high levels of digital health literacy. However, one-third of all students (4282/14,098, 30.4%) reported having problems finding the correct information on a particular health-related topic. Also, almost half of all students (5964/14,103, 42.3%) had problems evaluating the reliability of web-based information, which includes difficulties in identifying commercial interests behind the information presented in the news (5489/14,097, 38.9%). Moreover, the greatest challenges were related to assessing the reliability of COVID-19–related information and to judging whether commercial interests were attached to this information. Female students reported more difficulties in searching and evaluating web-based COVID-19–related information than male students.

Although digital health literacy levels were sufficient in a large proportion of the respondents, the results must be viewed in a more differentiated way. Germany applied a very successful health communication strategy [4] based on easily understandable and easy-to-use health information regarding COVID-19 (eg, washing hands, physical distancing, wearing masks), which was of low complexity compared to other health or disease information [23]. The communication mode was primarily push-based, directing information toward people through all media and communication channels. In comparison, for noncrisis communication, people must supply themselves with information (pull communication) to a greater extent, which requires active searching for information and hence requires strong health literacy. The way in which communication was altered (push vs pull) could explain the lack of differences in digital health literacy levels in relation to socioeconomic status, which are usually found in health literacy studies [22,37]. In addition, our study reports on the state of students' digital health literacy levels during the early stages of the so-called first wave of the pandemic, at a time when adherence to policies of measures to protect against COVID-19 was high. However, this could change in the current second wave, when people lose trust in official sources and the support for compliance with official recommendations diminishes. Declining support of public measures can already be observed in Germany. Demonstrations against restrictions are taking place, and people are demanding a return to prepandemic conditions and the reopening of the economy, all of which manifests through a refusal to apply the recommended protective measures (eg, no physical distancing, no face masks) [38].

Data protection and security is also an important issue in the context of digitalization and of the digital transformation of society. Using digital health services and communicating about health topics on the internet and on social media requires particular communication technologies to ensure user safety and user-friendliness. Our findings indicate that one-third of all students (4768/13,589, 35.1%) reported problems judging whether a third party can read their messages posted on the web. Studies on web-based data protection in the German population showed that 72% of respondents doubted the safety of the personal data they shared on the internet [39]. Moreover, 55% even believed that they had no control over what happens to their web-based data [40]. Despite its importance, we were required to exclude the subdimension "protecting privacy" from further bivariate analyses due to low reliability. When introducing the original DHLI, van der Vaart and Drossaert [20] also reported an unsatisfactory Cronbach alpha for this subscale (α=.57). Although this supports the validity of our study findings, it also suggests a need for further refinement (eg, by reformulating the item "…do you find it difficult to judge who can read along?“ to "…do you find it difficult to judge how the security of your private information is secured by the media provider?" to emphasize the role of protective measures taken by the media provider).

The most preferred sources to look for web-based health- and COVID-19–related information among students included search engines, news portals, and websites of public bodies, followed by social media and video portals. Favorite search topics were the current spread of COVID-19 (12,648/14,114, 89.6%), restrictions (12126/14,114, 85.9%), recommendations and risk assessments (10,975/14,114, 77.8%), and COVID-19 symptoms (10089/14,114, 71.5%). Similarly to earlier studies on population health literacy, which focus on both generic health literacy [22,41] and health literacy in relation to COVID-19 [15,23], making a judgment about the reliability of COVID-19–related information in the media and identifying potential commercial interests represent the most difficult tasks. There is also a significant positive association between having sufficient levels of digital health literacy and accessing more trustworthy and thus more reliable web-based health content. Students with higher levels of digital health literacy in the dimension of "evaluating information reliability" accessed the official websites of public bodies and agencies more often and turned less often to sources such as support communities, including forums, and social media compared to students with lower competencies in this dimension. While the ability to seek information and to produce and provide information did not show any significant differences across digital health literacy levels, students with higher abilities to determine the personal relevance of the information they obtained show similar patterns to those shown for the subscale "evaluating reliability."

Interestingly, only one-fifth of students (2921/14,114, 20.7%) reported having searched for information related to psychological stress and the consequences of the COVID-19 pandemic on mental health. This finding is surprising, as other studies show that the COVID-19 pandemic has enormous effects on mental health [42,43], and an infodemic can trigger an epidemic of fear and anxiety [28]. On the other hand, it should be emphasized that this survey was conducted at the beginning of the first wave of the pandemic and that psychological problems became more important as the pandemic progressed. Therefore, reliable and trustworthy (mental) health information is key in this situation for citizens to act upon information and knowledge provided by governments, health authorities, and scientists, and thereby to help slow the spread of COVID-19 [14,16,44-47]. In this context, infodemiology becomes important to better understand communication patterns, information routes and content, and how they affect behaviors, attitudes, and health status [28]. Citizen behavior, however, must be facilitated by adequate government actions and policies that provide not only health information but also health, social, and economic services for citizens to cope with the situation [14]. The impact of the ongoing COVID-19 infodemic places an additional burden on web-based health information seekers. This threat amplifies the negative effects of low digital health literacy. In their representative survey of COVID-19–related health literacy during the pandemic, Okan and colleagues [23] found that 56% of the German population felt confused about the vast amount of information regarding COVID-19. Women, younger age groups, and families with children younger than 18 years in their household are significantly more affected. At the same time, people with lower income and who reside in federal states of the former East Germany were found to feel less informed than their counterparts. This ongoing study highlights that the infodemic must be acknowledged "as a meta-risk in its own right" that aggravates the current situation [23]. Therefore, this infodemic requires particular attention during the COVID-19 emergency, which includes public policy strategies aiming to address the toxic spread of misinformation and disinformation about SARS-CoV-2 and COVID-19 [14,15,23]. Moreover, producers, providers, and suppliers of health information must ensure that information is evidence-based and adheres to health literacy principles, including barrier-free and easy access, user-friendliness and ease of understanding, cultural appropriateness, and relevance for everyday public use [21,23]. Social media platforms should also counteract the spread of misinformation and disinformation about COVID-19. The fight against misinformation and disinformation should become an important issue in public policy [17,18]. As proposed by Gunter Eysenbach, the four pillars to fight an infodemic include (1) infoveillance (the monitoring of information), (2) strengthening health literacy and digital health literacy in the population, (3) applying constant knowledge refinement (eg, fact checking), and (4) adequate knowledge transfer and minimizing political and commercial influence on health information [45]. This is supported by the World Health Organization within their infodemic management framework, which suggests six policy recommendations to manage infodemics during an emergency such as the COVID-19 pandemic [46]. These recommendations include (1) basing interventions and messages on the latest evidence, (2) applying knowledge transfer and making health information easy to understand, (3) collaborating with communities to better understand their information needs, (4) analyzing information impact and cooperating with social media platforms, information suppliers, and civil society, (5) informing these actions by reliable information and adapt action based on the respective and latest narratives, and (6) further improving infodemic management by all means necessary and also through interdisciplinary research collaboration [46]. Skills to navigate digital information environments were already crucial before the COVID-19 pandemic to mitigate the effects of digital inequalities [47,48]. These skills have become even more essential during the pandemic, as the importance and use of communication technologies and media have changed massively since the outbreak of COVID-19 [14,16,45,48,49].

The most important finding of the stratified analyses is that among students with limited digital health literacy, female students reported having more problems finding the correct information and evaluating the reliability of COVID-19–related information. In Germany, women often have more care responsibilities and are generally more engaged with health issues than men [23], and they are also more active in searching the internet for health information [49,50]. This may lead them to be more critical vis-à-vis health information on COVID-19, as they have a more sensible awareness that not all information is reliable. In addition to this, a recent study showed that women are much more worried about the sheer amount of COVID-19–related information on the internet [23]. They were more concerned when they had children ≤18 years of age. Many young women are faced with difficulties and challenges when they search for and evaluate health information, especially because there is so much conflicting information on COVID-19 available on the web.

To sum up, the findings from this study raise concern and have important implications for public health. First, problems related to access to accurate and situation-specific information in the context of a public health emergency may lead to the use of invalid information, which is unhelpful or even detrimental to the causes of slowing infection rates and sustaining a successful infectious disease strategy. Second, when students access disinformation or false information and they have difficulty making judgments about the correctness of the information, they will most likely not identify that information for what it is (eg, “fake news,” commercial messaging). In turn, using the wrong information can again cause harm and impede engagement in effective health behaviors. Third, feeling safe in the digital world, especially when seeking health-related information and interacting with others about health concerns, is a critical issue. Many students expressed uncertainties regarding the safety of personal information shared on the web. These findings suggest the need to implement health education measures to strengthen students’ health literacy capacities. In addition, there is a need for more accurate public health information platforms to provide timely and evidence-based information with a view to inform individual behavior and system-level responses. Studies on health literacy in Germany conducted in 2011 [37], 2013 [41,50], and 2014 [22,41] have shown that half of the adult population, including both younger and older adults, have limited health literacy. In response, health literacy policy initiatives were launched, such as the science and civil society–led German National Action Plan on Health Literacy [51] and the Alliance for Health Literacy [52], which is led by the federal Ministry of Health. These initiatives focus on strengthening population health literacy, starting in early childhood and at school, to enable children to grow into health-literate adults. However, little progress has been made since then, and a health education curriculum that addresses health literacy is still lacking in Germany. Furthermore, people with lower education in Germany have more often lower levels of health literacy [22,37]. If students, who belong to a population group with higher education, already have difficulties with their digital health literacy, it can be assumed that people with less education are also vulnerable to having lower levels of health literacy and associated information tasks, such as finding, understanding, and evaluating COVID-19–related information on the web.

Our study has several limitations. The sample, although weighted, is not representative of all university students in Germany. We may have missed many students who use the internet less frequently and those who may have been troubled due to university closures and associated changes to their lives. The implications may not be transferable to other populations and age groups in Germany. Additionally, students in Germany are privileged in terms of educational achievement and therefore in terms of socioeconomic status compared to people seeking a tertiary education with non–degree level requirements. This survey was conducted in the early days of the first wave of the pandemic, when adherence was high; this could explain the finding that students perceived information tasks to be easy to undertake and therefore reported high levels of digital health literacy. This may not be the case in a second phase, after enduring lengthy restrictions on everyday and university life activities and rapidly emerging conflicting information on COVID-19, all of which could make judgment much more difficult. Due to the effects of the COVID-19 pandemic on physical contact and face-to-face meeting, we had to use a web-based survey in adherence to German COVID-19 policies, whereas the developers of the original questionnaire, van der Vaart and Drossaert [20], highlighted that the application of a web-based questionnaire may exclude people with weak digital competencies. Therefore, a potential bias in our sample is that it may have excluded students who use the internet to a lesser extent or those with lower digital competencies. Nevertheless, due to web-based activities related to their studies (eg, access to e-learning and university communication platforms) and given that most German universities provide their services via web-based systems, students in Germany in general represent a proportion of the population who have more intersections with the digital world, inevitably use the internet more often, and have a higher affinity to using web-based media content.

Our findings show that overall, the level of digital health literacy in relation to dealing with web-based COVID-19–related information was high. However, a significant proportion of university students still face difficulties with certain abilities to deal with information, such as finding the right information and evaluating its reliability. There is a need to strengthen the digital health literacy capacities of university students, particularly female students, using tailored interventions. Actions must also include the design of interventions to increase the quality of health information on the internet, to implement fact-checking strategies in web-based and social media, and to increase the health literacy of people who produce, supply, and provide health information and services on the web. For example, universities can provide courses on digital health literacy and health information to their staff and students and can also disseminate reliable news on COVID-19 through their web-based channels. Raising awareness among universities and education administrators might aid the emergency response, and it could also increase the health literacy responsiveness of organizations and students. The benefit of the COVID-HL survey is that it provides first-time knowledge that could help decision-makers develop policies and programs that foster healthy and protective behaviors, plan for preventive measures, and promote adherence to COVID-19 policies, on the basis of students' needs in terms of digital health literacy. Digital health literacy will empower university students and all other population groups to take greater control in the prevention and spread of COVID-19, which in turn is likely to lead to better health outcomes.

Authors' Contributions

KD and OO contributed equally and prepared the first draft of the manuscript. MM, AYML, RR, ED, and KR provided feedback on the first draft. KD, OO, and KR designed and conducted the survey. MM, AYML, RR, ED, and KR helped with the analyses and contributed to data interpretation. All authors contributed to writing and revising the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of scales and items.
[DOCX File , 97 KB - jmir_v23i1e24097_app1.docx]

References

1. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020 Feb 22;395(10224):565-574 [FREE Full text] [doi: 10.1016/S0140-6736(20)30251-8] [Medline: 32007145]
2. World Health Organization. Novel Coronavirus (2019-nCoV) - Situation Report 13. 2020 Feb 02. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf [accessed 2020-09-05]
3. European Centre for Disease Prevention and Control. ECDC Communicable Disease Threats Report (CDTR). URL: https://www.ecdc.europa.eu/sites/default/files/documents/Communicable-disease-threats-report-21-mar-2020-PUBLIC.pdf [accessed 2020-09-05]
4. Wieler L, Gottschalk R. Emerging COVID-19 success story: Germany?s strong enabling environment. Our World in Data. 2020 Jun 30. URL: https://ourworldindata.org/covid-exemplar-germany [accessed 2020-09-05]
5. Robert Koch Institut. Ergänzung zum Nationalen Pandemieplan – COVID-19 – neuartige Coronaviruserkrankung. 2020 Mar 04. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Ergaenzung_Pandemieplan_Covid.html [accessed 2020-09-05]
6. Bundesministerium für Gensundheit. Zusammen gegen Corona. Aktuelle Informationen zum neuartigen Coronavirus/Covid-19 Together against corona. URL: https://www.zusammengegencorona.de/informieren [accessed 2020-09-05]
7. Bundesministerium für Gesundheit. Coronavirus in Deutschland. URL: https://www.bundesregierung.de/breg-de/themen/coronavirus [accessed 2020-09-05]
8. Bundeszentrale für gesundheitliche Aufklärung. Informationen zum neuartigen Coronavirus / COVID-19 Information about the novel coronavirus/COVID-19, webpage on German. URL: https://www.infektionsschutz.de [accessed 2020-09-05]

9. Robert Koch-Institut. COVID-19 (Coronavirus SARS-CoV-2). Webpage in German. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/nCoV.html [accessed 2020-09-05]
10. Patienten-Information.de. Corona – verlässliche Informationen zu COVID 19. URL: https://www.patienten-information.de/kurzinformationen/coronavirus [accessed 2020-09-05]
11. Zweites Deutsches Fernsehen. Coronavirus: Alles zu Covid-19. URL: https://www.zdf.de/nachrichten/thema/coronavirus-covid-19-100.html [accessed 2020-09-05]
12. Norddeutscher Rundfunk. Coronavirus-Update mit Christian Drosten. URL: https://www.ndr.de/nachrichten/info/podcast4684.html [accessed 2020-09-05]
13. Corona-Krise. Das Erste. URL: https://www.daserste.de/information/politik-weltgeschehen/coronakrise/index.html [accessed 2020-09-05]
14. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health 2020 May;5(5):e249-e250 [FREE Full text] [doi: 10.1016/S2468-2667(20)30086-4] [Medline: 32302535]
15. Okan O, de Sombre S, Hurrelmann K, Berens EM, Bauer U, Schaeffer D. Covid-19-Gesundheitskompetenz der Bevölkerung COVID-19 based health literacy in the German population. Monitor Versorgungsforschung 2020 Mar;13:40-45 [FREE Full text]
16. Okan O, Sørensen K, Messer M. COVID-19: A guide to good practice on keeping people well informed. The Conversation. 2020 Mar 19. URL: https://theconversation.com/covid-19-a-guide-to-good-practice-on-keeping-people-well-informed-134046 [accessed 2020-09-05]
17. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
18. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
19. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
20. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
21. Sørensen K. Covid-19: Digital health literacy is a key to saving time, costs and lives. ICT & health. 2020 Mar 30. URL: https://www.ictandhealth.com/news/covid-19-digital-health-literacy-is-a-key-to-saving-time-costs-and-lives [accessed 2020-09-05]
22. Schaeffer D, Berens E, Vogt D. Health Literacy in the German Population. Dtsch Arztebl Int 2017 Jan 27;114(4):53-60 [FREE Full text] [doi: 10.3238/arztebl.2017.0053] [Medline: 28211318]
23. Okan O, Bollweg TM, Berens E, Hurrelmann K, Bauer U, Schaeffer D. Coronavirus-Related Health Literacy: A Cross-Sectional Study in Adults during the COVID-19 Infodemic in Germany. Int J Environ Res Public Health 2020 Jul 30;17(15):5503 [FREE Full text] [doi: 10.3390/ijerph17155503] [Medline: 32751484]
24. Van den Broucke S. Why health promotion matters to the COVID-19 pandemic, and vice versa. Health Promot Int 2020 Apr 01;35(2):181-186 [FREE Full text] [doi: 10.1093/heapro/daaa042] [Medline: 32297931]
25. Horgan Á, Sweeney J. University students' online habits and their use of the Internet for health information. Comput Inform Nurs 2012 Aug;30(8):402-408. [doi: 10.1097/NXN.0b013e3182510703] [Medline: 22525044]
26. Dadaczynski K, Okan O, Messer M, Rathmann K. Digitale Gesundheitskompetenz von Studierenden in Deutschland während der Corona-Pandemie. Ergebnisse einer bundesweiten Online-Befragung. Hochschule Fulda. 2020 Aug 14. URL: https://fuldok.hs-fulda.de/opus4/843 [accessed 2020-09-05]
27. Nguyen HT, Do BN, Pham KM, Kim GB, Dam HTB, Nguyen TT, et al. Fear of COVID-19 Scale-Associations of Its Scores with Health Literacy and Health-Related Behaviors among Medical Students. Int J Environ Res Public Health 2020 Jun 11;17(11):4164 [FREE Full text] [doi: 10.3390/ijerph17114164] [Medline: 32545240]
28. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
29. Statistische Daten zu Studienangeboten an Hochschulen in Deutschland. Studiengänge, Studierende, Absolventinnen und Absolventen. Wintersemester 2019. In: Hochschulrektorenkonferenz. Berlin, Germany: HRK; 2019.
30. Robert Koch-Institut: COVID-19-Dashboard. Webpage in German. Robert Koch-Institut. URL: https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4 [accessed 2020-09-05]
31. EFS Survey, Version EFS Winter 2018. Questback GmbH. Cologne: Questback GmbH; 2019. URL: https://community.questback.com/s/questback-efs-survey [accessed 2021-01-09]
32. Hoebel J, Müters S, Kuntz B, Lange C, Lampert T. Measuring subjective social status in health research with a German version of the MacArthur Scale. Article in German. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Jul;58(7):749-757. [doi: 10.1007/s00103-015-2166-x] [Medline: 25986532]

33. Schricker J, Rathmann K, Dadaczynski K. Soziale Unterschiede in der Gesundheitskompetenz von Studierenden: Ergebnisse einer Online-Studie an der Technischen Universität Dortmund. Präv Gesundheitsf 2019 Jul 31;15(1):8-14. [doi: 10.1007/s11553-019-00731-6]
34. Marstedt G. Das Internet: Auch Ihr Ratgeber für Gesundheitsfragen? Bevölkerungsumfrage zur Suche von Gesundheitsinformationen im Internet und zur Reaktion der Ärzte. Gütersloh, Germany: Bertelsmann Stiftung; 2018.
35. Die Datenbank des Statistischen Bundesamtes. Federal Statistical Office of Germany. URL: https://www-genesis.destatis.de/genesis/online [accessed 2021-01-07]
36. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hoboken, NJ: Taylor and Francis; 2013.
37. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
38. Sueddeutsche Z. Berlin: 20.000 Menschen demonstrieren gegen Corona-Auflagen. Süddeutsche Zeitung. 2020 Aug 01. URL: https://www.sueddeutsche.de/politik/demonstrationen-berlin-20-000-menschen-demonstrieren-gegen-corona-auflagen-dpa.urn-newsml-dpa-com-20090101-200801-99-04333 [accessed 2020-09-05]
39. Nagler K, Schröder D, Schilling A, Weber T. DsiN-Sicherheitsindex 2020. Studie von Deutschland sicher im Netz. 2020. URL: https://www.sicher-im-netz.de/dsin-sicherheitsindex-2020 [accessed 2020-09-05]
40. Studie zu Datenschutz: Mehrheit der Deutschen zweifelt an Datensicherheit. Sinus Institut. URL: https://www.sinus-institut.de/veroeffentlichungen/meldungen/detail/news/studie-zu-datenschutz-mehrheit-der-deutschen-zweifelt-an-datensicherheit/news-a/show/news-c/NewsItem [accessed 2020-09-05]
41. Berens E, Vogt D, Messer M, Hurrelmann K, Schaeffer D. Health literacy among different age groups in Germany: results of a cross-sectional survey. BMC Public Health 2016 Nov 09;16(1):1151 [FREE Full text] [doi: 10.1186/s12889-016-3810-6] [Medline: 27829401]
42. Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: Development and Initial Validation. Int J Ment Health Addict 2020 Mar 27:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00270-8] [Medline: 32226353]
43. Satici B, Gocet-Tekin E, Deniz ME, Satici SA. Adaptation of the Fear of COVID-19 Scale: Its Association with Psychological Distress and Life Satisfaction in Turkey. Int J Ment Health Addict 2020 May 08:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00294-0] [Medline: 32395095]
44. Sentell T, Vamos S, Okan O. Interdisciplinary Perspectives on Health Literacy Research Around the World: More Important Than Ever in a Time of COVID-19. Int J Environ Res Public Health 2020 Apr 26;17(9):3010 [FREE Full text] [doi: 10.3390/ijerph17093010] [Medline: 32357457]
45. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
46. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
47. Beaunoyer E, Dupéré S, Guitton MJ. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. Comput Human Behav 2020 Oct;111:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]
48. Guitton MJ. Cyberpsychology research and COVID-19. Comput Human Behav 2020 Oct;111:106357 [FREE Full text] [doi: 10.1016/j.chb.2020.106357] [Medline: 32362721]
49. Bidmon S, Terlutter R. Gender Differences in Searching for Health Information on the Internet and the Virtual Patient-Physician Relationship in Germany: Exploratory Results on How Men and Women Differ and Why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
50. Jordan S, Hoebel J. [Health literacy of adults in Germany: Findings from the German Health Update (GEDA) study]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Sep;58(9):942-950. [doi: 10.1007/s00103-015-2200-z] [Medline: 26227894]
51. Schaeffer D, Hurrelmann K, Bauer U, Kolpatzik K. National Action Plan Health Literacy. Promoting Health Literacy in Germany. Berlin, Germany: KomPart; 2018. URL: https://www.nap-gesundheitskompetenz.de/app/download/7775450763/National%20Action%20Plan%20Health%20Literacy.pdf?t=1607442028 [accessed 2021-01-08]
52. Bundesministerium für Gesundheit. Allianz für Gesundheitskompetenz. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/E/Erklaerungen/Allianz_fuer_Gesundheitskompetenz_Abschlusserklaerung.pdf [accessed 2020-09-05]

Abbreviations

DHLI: Digital Health Literacy Instrument
SSS: subjective social status

Edited by G Eysenbach; submitted 05.09.20; peer-reviewed by E Sakellari, S Bidmon; comments to author 13.10.20; revised version received 30.10.20; accepted 08.12.20; published 15.01.21.

Please cite as:
Dadaczynski K, Okan O, Messer M, Leung AYM, Rosário R, Darlington E, Rathmann K
Digital Health Literacy and Web-Based Information-Seeking Behaviors of University Students in Germany During the COVID-19 Pandemic: Cross-sectional Survey Study
J Med Internet Res 2021;23(1):e24097
URL: http://www.jmir.org/2021/1/e24097/
doi:10.2196/24097
PMID:33395396

Original Paper

Asynchronous Distance Learning of the National Institutes of Health Stroke Scale During the COVID-19 Pandemic (E-Learning vs Video): Randomized Controlled Trial

Mélanie Suppan[1], MD; Loric Stuby[2], CAS; Emmanuel Carrera[3], MD; Philippe Cottet[4], MD; Avinash Koka[4], MD; Frédéric Assal[5], MD; Georges Louis Savoldelli[1,6], MD, MEd; Laurent Suppan[4], MD

[1]Division of Anaesthesiology, Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals and Faculty of Medicine, Geneva, Switzerland

[2]Genève TEAM Ambulances, Geneva, Switzerland

[3]Stroke Center, Department of Neurology, Geneva University Hospitals and Faculty of Medicine University of Geneva, Geneva, Switzerland

[4]Division of Emergency Medicine, Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals and Faculty of Medicine, Geneva, Switzerland

[5]Cognitive Neurology Unit, Neurology, Department of Clinical Neuroscience, Geneva University Hospital and Faculty of Medicine, Geneva, Switzerland

[6]Unit of Development and Research in Medical Education, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:
Laurent Suppan, MD
Division of Emergency Medicine
Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
Geneva University Hospitals and Faculty of Medicine
Rue Gabrielle-Perret-Gentil 4
Geneva, 1211
Switzerland
Phone: 41 795532579
Email: laurent.suppan@hcuge.ch

Abstract

Background: The COVID-19 pandemic has considerably altered the regular medical education curriculum while increasing the need for health care professionals. Senior medical students are being incrementally deployed to the front line to address the shortage of certified physicians. These students, some of whom will be fast-tracked as physicians, may lack knowledge regarding the initial management of time-critical emergencies such as stroke.

Objective: Our aim was to determine whether an e-learning module could improve asynchronous distance knowledge acquisition of the National Institutes of Health Stroke Scale (NIHSS) in senior medical students compared to the traditional didactic video.

Methods: A randomized, data analyst–blinded web-based trial was conducted at the University of Geneva Faculty of Medicine between April and June 2020. Fifth year medical students followed a distance learning path designed to teach the NIHSS. The control group followed the traditional didactic video created by Patrick Lyden, while the e-learning group followed the updated version of a previously tested, highly interactive e-learning module. The main outcome was the score on a 50-question quiz displayed upon completion of the learning material. The difference in the proportion of correct answers for each specific NIHSS item was also assessed.

Results: Out of 158 potential participants, 88 started their allocated learning path and 75 completed the trial. Participants who followed the e-learning module performed better than those who followed the video (38 correct answers, 95% CI 37-39, vs 35 correct answers, 95% CI 34-36, $P<.001$). Participants in the e-learning group scored better on five elements than the video group: key NIHSS concepts ($P=.02$), the consciousness – global item ($P<.001$), the facial palsy item ($P=.04$), the ataxia item ($P=.03$), and the sensory item ($P=.04$).

Conclusions: Compared to the traditional didactic video, a highly interactive e-learning module enhances asynchronous distance learning and NIHSS knowledge acquisition in senior medical students.

(J Med Internet Res 2021;23(1):e23594) doi:10.2196/23594

KEYWORDS

stroke; COVID-19; e-learning; medical student; medical education; online learning; randomized controlled trial; video

Introduction

The swift strengthening of public health policies in the context of the COVID-19 crisis has wrought deep changes in the regular medical education curricula of many countries [1-4] while also increasing the need for health care professionals, including physicians. Senior medical students are being incrementally used on the front lines to address the shortage of these professionals [5,6], and other students may soon be required to follow suit [7]. Accelerated graduation procedures have also been described in some regions [8]. Senior medical students as well as some of these fast-tracked physicians may lack knowledge regarding the initial management of specific emergencies such as stroke. Stroke is a time-critical emergency that must be treated swiftly to improve functional and vital prognoses [9]; however, disruptions in acute stroke pathways have been described in the wake of the COVID-19 pandemic [10]. The National Institutes of Health Stroke Scale (NIHSS) is widely used to assess stroke victims [11], and senior medical students as well as junior residents should be familiar with its application.

Traditional classroom or bedside teaching can be difficult to conduct in certain situations, such as a pandemic [12-14]. Many universities have strived to increase distance learning capabilities, thereby highlighting the potential benefits of electronic learning (e-learning) [15-17]. E-learning is a generic term that includes many types of technologically enhanced learning materials [18-20]. Asynchronous distance learning using these methods has yielded mixed results, probably due to differences in the quality of the content and the mode of delivery [21].

Since the release of Patrick Lyden’s didactic video in 1994 [22], the development of NIHSS teaching material has been rather limited. We have recently shown that compared to this didactic video, a highly interactive e-learning module improved NIHSS knowledge acquisition in paramedics [23]. We defined this module as “highly interactive” because it uses multiple learning mechanics to promote interaction and engagement. Among these mechanics, preventing content skipping [24] and providing feedback tailored to the user's answer were the most prominent [25]. Branching logic was extensively used to create this feedback.

This first study was performed with the participants present at the study site; therefore, they could immediately access technical support if needed. Moreover, although most results favored the use of the e-learning module, the control group was better at scoring the ataxia element than the e-learning group. Although video extracts were used within the e-learning module to demonstrate the assessment of almost all NIHSS items, the chapter regarding the ataxia element did not contain any video extracts. We therefore hypothesized that systematically embedding videos could improve NIHSS learning acquisition, and we updated the module accordingly.

Given the need for social distancing during the COVID-19 pandemic, our goal was to compare medical students’ asynchronous distance learning of the NIHSS using two different teaching tools: the gold standard didactic video and the updated version of our e-learning module.

Methods

Study Design and Setting

We performed a randomized, controlled, data analyst–blinded, web-based trial following the CONSORT-EHEALTH guidelines and incorporating relevant elements from the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist [26,27]. The study took place between April 28 and June 8, 2020, in Geneva, Switzerland. Fifth year medical students at the University of Geneva Faculty of Medicine (UGFM) were invited to take part in this trial on a voluntary basis.

Standard Protocol Approvals, Registrations, and Participant Consent

Because the study included no patients and as no health outcomes were recorded, trial registration was not required according to the International Committee of Medical Journal Editors guidelines. Although the participants were not part of a vulnerable group according to Swiss federal law on human research [28], we filed a jurisdictional enquiry, and the regional ethics committee issued a “Declaration of no objection” (Req 2020-00474). The study was also approved by the Board of the Teaching Committee of the UGFM. Informed consent was gathered electronically.

Enrollment

After gathering the necessary authorizations, the UGFM students’ secretary transmitted the exact number of fifth year medical students to MS, who performed a 1:1 computer assisted randomization without having access to any other data regarding the students. MS then created specific identifiers that were transmitted back to the UGFM along with a mailing template Multimedia Appendix 1. The UGFM staff were therefore unable to determine the students’ allocations or results. In addition, we were prevented from determining the students’ identities.

The students were informed of the goals of the investigation, were given information regarding data security and anonymization procedures, and were supplied with the email addresses of three investigators to allow them to ask further questions. Students who elected to browse the website were provided with additional information as well as with a link to a full 4-page consent form in PDF format that they could either print or save. Using their identifiers to log into the site was considered as acceptance to participate in the study. All participants were free to withdraw at any time. No financial incentive was provided.

Web-Based Platform and Learning Material

We created a specific web-based platform under the Joomla 3.9 content management system (Open Source Matters) [29]. The control was Patrick Lyden's original video, which was subtitled in French [22]. The experiment used version 21 of our e-learning module, which was developed using Articulate Storyline 3 (Articulate Global). This software enables the creation of many types of interactive content, including gamified modules and serious games [30,31], which can be accessed on regular computers as well as on smartphones and tablets.

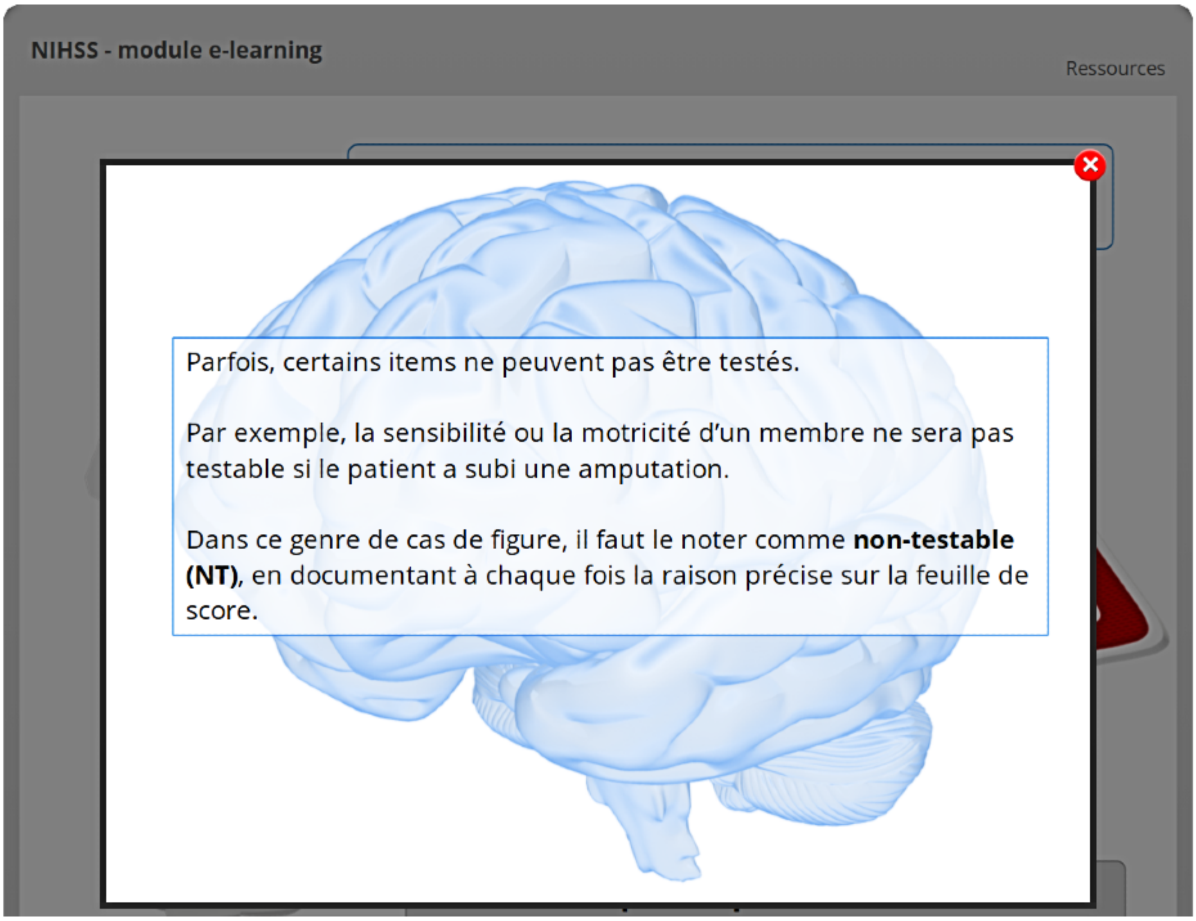
The e-learning module contains 16 independent chapters. The first chapter is the introduction, which is automatically displayed when the module is launched. Prevention of content skipping is the first learning mechanic used in the module, and it already appears in the introductory slides (Figures 1-3) [24].

A table of contents is displayed as soon as the user has completed the introduction. The user can then choose to review the introduction or to access any other chapter apart from the summary (Figure 4).

Figure 1. Prevention of content skipping. The user cannot click on the *Cliquez ici pour continuer* (Click here to continue) button until both blue buttons have been clicked.

Pour certains items, il faudra tenir compte de subtilités ! Cliquez sur les boutons pour les découvrir.

Figure 2. The user has clicked on one of the two buttons, and the learning content is now displayed in a lightbox. The *Cliquez ici pour continuer* (Click here to continue) button, which is slightly visible in the background, is still grey; therefore, it is inactive.

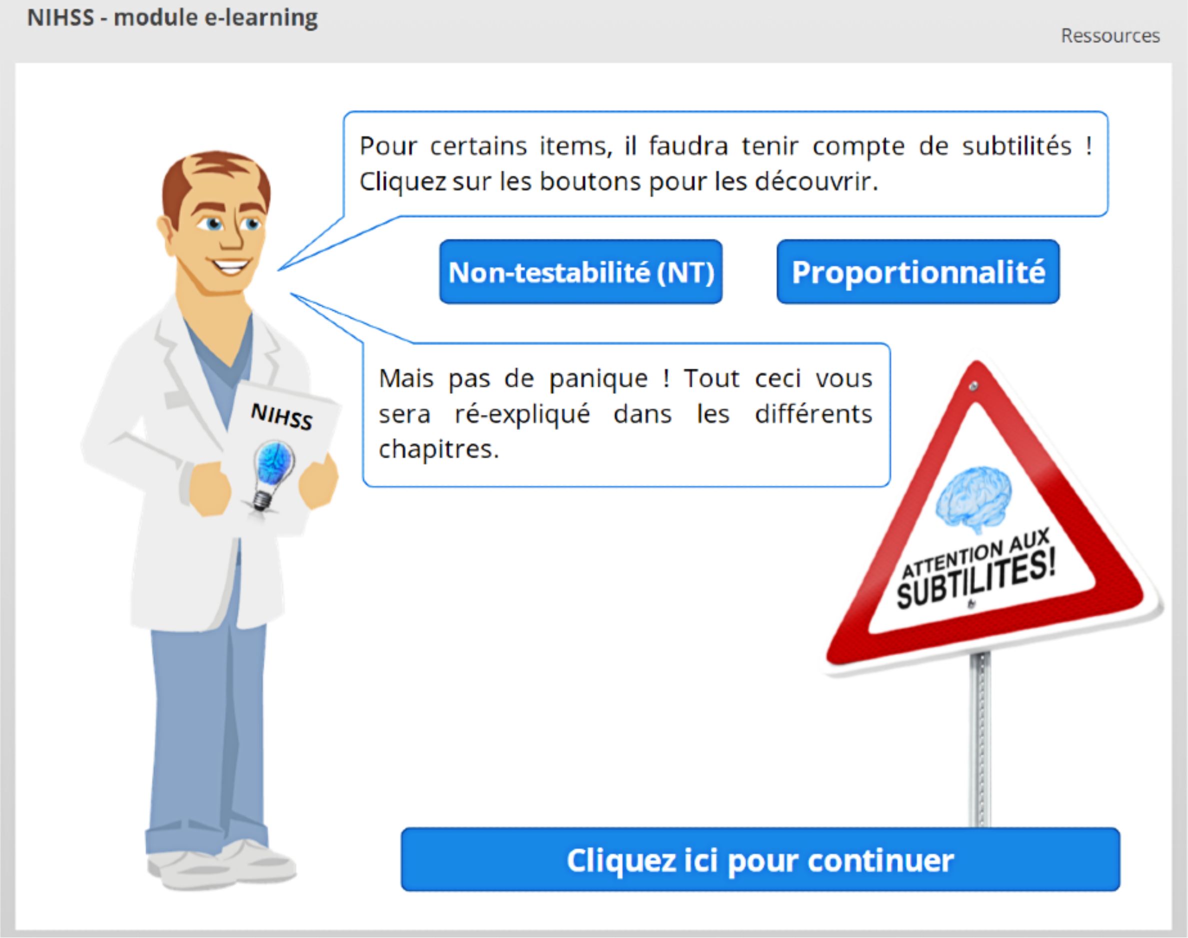


Parfois, certains items ne peuvent pas être testés.

Par exemple, la sensibilité ou la motricité d'un membre ne sera pas testable si le patient a subi une amputation.

Dans ce genre de cas de figure, il faut le noter comme **non-testable (NT)**, en documentant à chaque fois la raison précise sur la feuille de score.

Figure 3. Both buttons have been clicked, and the user has seen both lightboxed slides. The *Cliquez ici pour continuer* (Click here to continue) button has thus been activated and is now colored blue.



Pour certains items, il faudra tenir compte de subtilités ! Cliquez sur les boutons pour les découvrir.

Mais pas de panique ! Tout ceci vous sera ré-expliqué dans les différents chapitres.

Figure 4. Table of contents of the e-learning module. The *Résumé* (Summary) can only be accessed once all the other chapters have been completed.

NIHSS - module e-learning

Ressources

Introduction

1A Etat de conscience – global
1B Etat de conscience – questions
1C Etat de conscience – consignes
2 Oculomotricité
3 Champ visuel
4 Paralysie faciale
5 Motricité des bras
6 Motricité des jambes
7 Ataxie
8 Sensibilité
9 Langage
10 Dysarthrie
11 Extinction et négligence
Score pour le coma

Résumé

Thirteen chapters are used to explain each specific NIHSS item (3 chapters are used to cover the first item, “consciousness,” which is divided into 3 elements). Although the chapters are numbered consecutively according to the NIHSS scoring logic, the user can freely elect the order in which to follow the chapters.

All chapters include at least two learning mechanics. First, each chapter begins by displaying the NIHSS score specific to the scoring item, and users are once again prevented from skipping content, as they must click on each numbered button to discover the score (Figure 5).

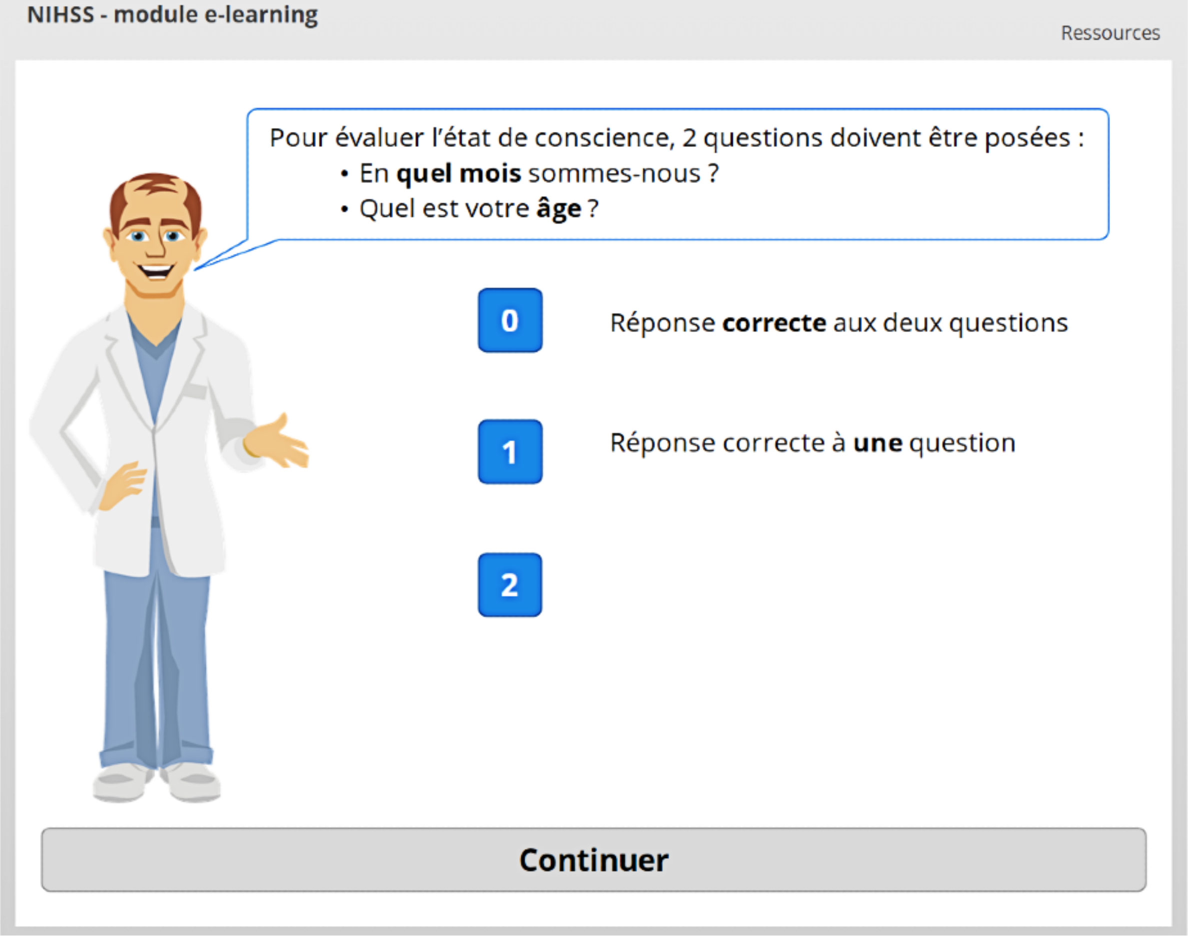
The second learning mechanic is linked to the use of subtitled videos. Video extracts are shown to the user, who must correctly score the NIHSS item (Figure 6). This version of the module contains video extracts in all chapters, including for the items related to dysarthria, level of consciousness – global, and level of consciousness – questions. In the previously studied version of our module (version 20), there were no video extracts for these items [23].

Feedback is provided for each question [25]: if the answer is incorrect, a clue is given (Figure 7), and the user has the opportunity to review the NIHSS item scoring.

If the answer was correct, feedback is also given to reinforce the message (Figure 8).

Specific interactions were designed to further illustrate particular elements, such as visual field deficits (Figure 9) or extinction and inattention (Figure 10).

Figure 5. Prevention of content skipping. The user cannot click on the *Continuer* (Continue) button until all the blue buttons have been clicked.



Pour évaluer l'état de conscience, 2 questions doivent être posées :

- En **quel mois** sommes-nous ?
- Quel est votre **âge** ?

0 Réponse **correcte** aux deux questions

1 Réponse correcte à **une** question

2

Figure 6. Video-based question. The user must choose the correct score for the patient displayed in the video before clicking on *Valider la réponse* (Validate the answer).

Quel est le niveau de conscience de ce patient?

Oui, un peu.

Figure 7. Wrong answer feedback and clue. The user can choose either to try again (*Réessayer*) or to review the scoring specific to this item by clicking *Cliquez ici pour afficher l'échelle* (Click here to display the scale).

NIHSS - module e-learning

Ressources

Quel est le niveau de conscience de ce patient?

Mauvaise réponse

N'avez-vous pas l'impression que ce patient a un excellent état de conscience?

○ 0 ● 1 ○ 2 ○ 3

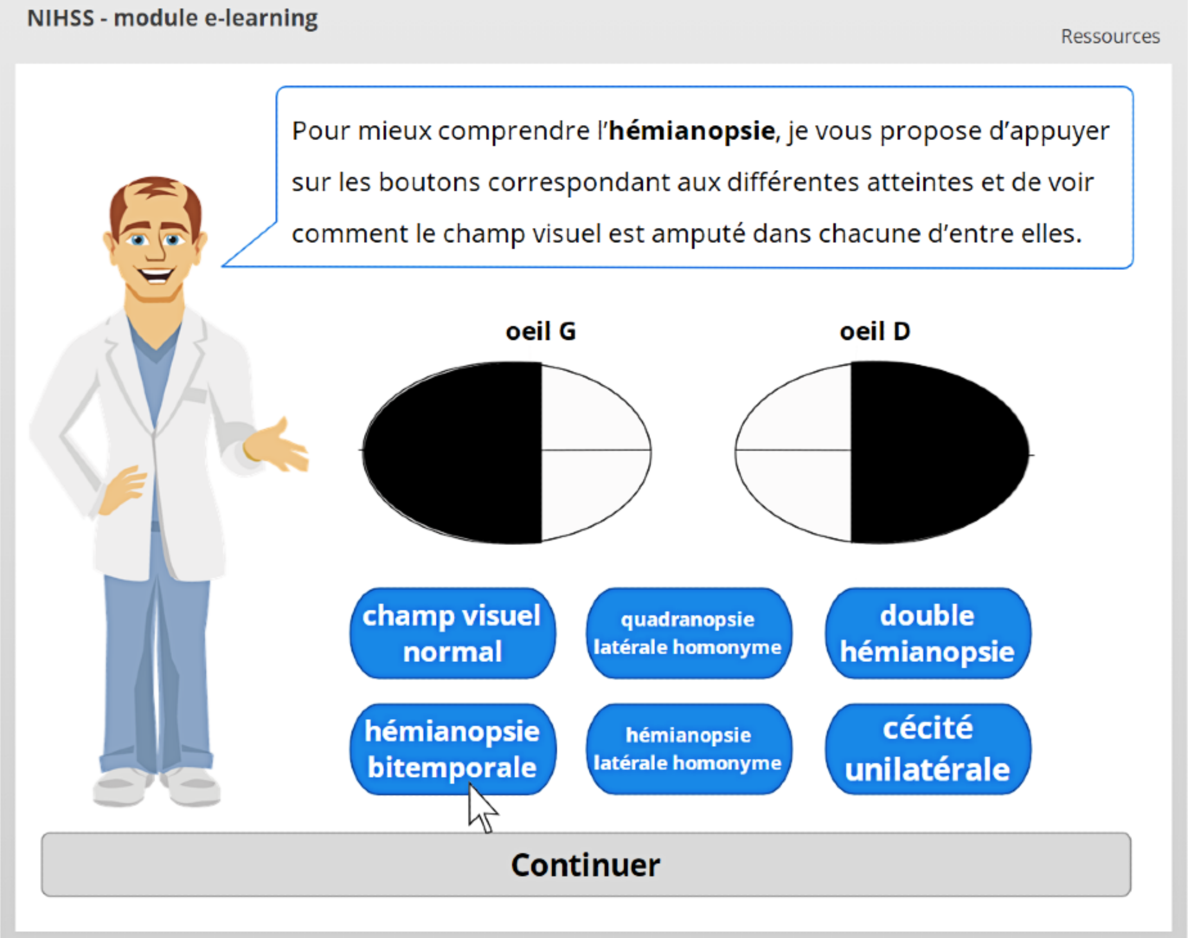
Figure 8. Positive feedback after a correct answer.

Quel est le niveau de conscience de ce patient?

Bonne réponse

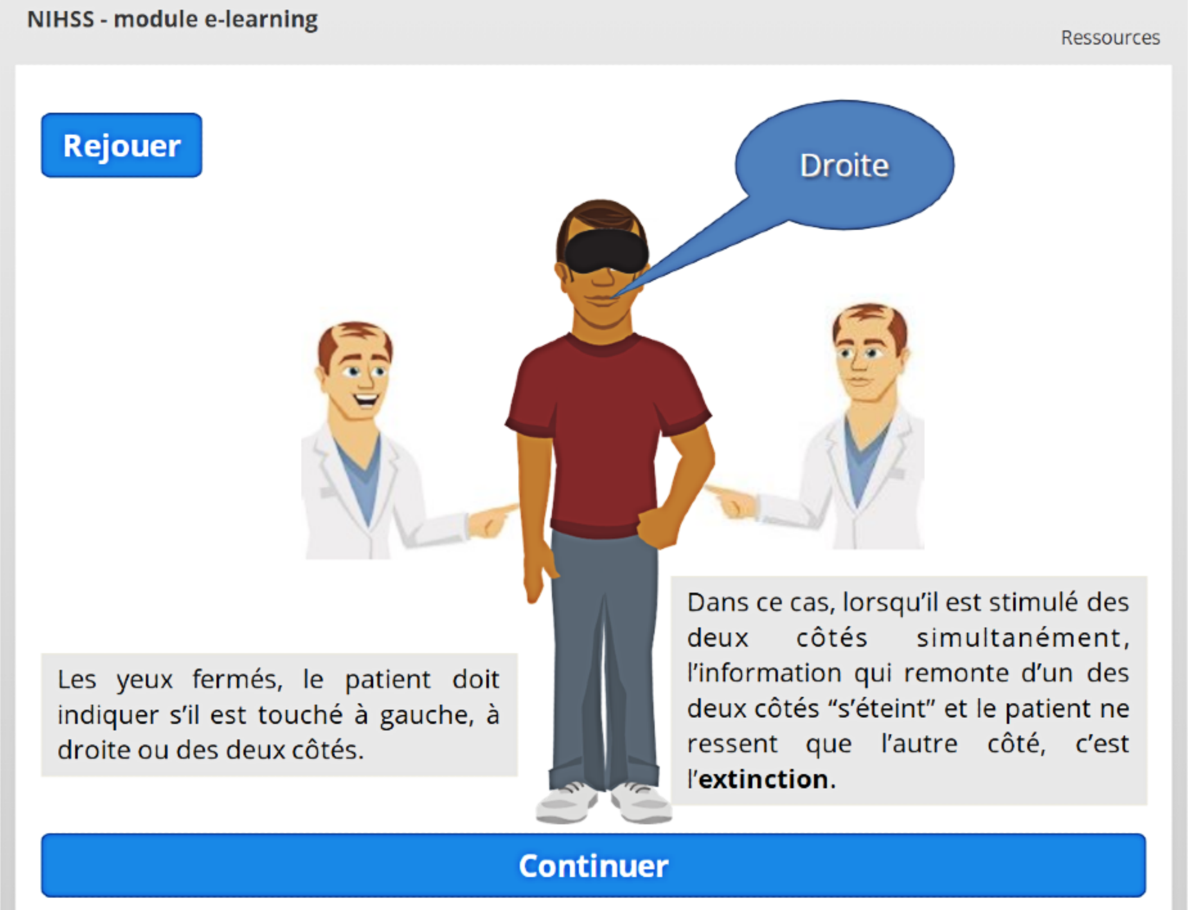
Bravo ! Vous avez sélectionné la bonne réponse. Ce patient est en effet alerte et a un état de conscience tout à fait normal.

Figure 9. Interactive explanation of visual field deficit.



Pour mieux comprendre l'**hémianopsie**, je vous propose d'appuyer sur les boutons correspondant aux différentes atteintes et de voir comment le champ visuel est amputé dans chacune d'entre elles.

Figure 10. Animation used to explain extinction and inattention.



Les yeux fermés, le patient doit indiquer s'il est touché à gauche, à droite ou des deux côtés.

Dans ce cas, lorsqu'il est stimulé des deux côtés simultanément, l'information qui remonte d'un des deux côtés "s'éteint" et le patient ne ressent que l'autre côté, c'est l'**extinction**.

The user can choose to exit the module at any time, as a prompt will be shown to allow the user to either resume the module or reset it. Before the summary chapter can be activated, users must complete a 14th chapter, which details the "coma score."

This e-learning module, along with its previously studied iteration, can be accessed freely on the internet [32].

Study Sequence

Immediately after login, the medical students learned which group they had been allocated to and were asked a first set of 6 questions displayed over a single page. Upon answering these questions, they could then access the learning material. No time limit was applied apart from the study end date (June 8, 2020). Once the learning material was completed, students were allowed to proceed to a 50-question quiz. This quiz was identical for all participants and contained five questions related to basic NIHSS concepts, followed by the clinical evaluation of 3 patients taken from Patrick Lyden's certification videos. The NIHSS elements were displayed and scored in sequence according to the NIHSS scoring logic. After finalizing the quiz, participants were given their overall score as well as the possibility to review all questions at will and were shown their answers along with the correct answers. Then, 4 questions, based on a 5-point Likert scale, were asked to assess secondary outcomes, such as satisfaction. Students were finally given access to both the video and the e-learning module to discover the other teaching modality and/or to review the one they had just followed.

Outcomes

The primary outcome of the study was the proportion of correct answers to the 50-question quiz. Secondary outcomes were the proportion of correct answers for each specific NIHSS item, user satisfaction, perceived adequacy of the time needed to complete the course, perceived difficulty of the course, probability that the participant would recommend the course, and whether the learning path had been completed over multiple days.

Data Collection and Curation

Data were securely stored on an encrypted MariaDB 5.5.5 database (MariaDB Foundation) located on a Swiss server before being extracted in comma-separated values (CSV) format. We used STATA (StataCorp) for data curation and anonymization.

Statistical Analysis

STATA 15.1 was used by L Stuby for statistical analysis. Incomplete answers to the 50-question quiz were not analyzed.

Normality was assessed by graphical evaluation and, if in doubt, we used the Shapiro-Wilk test. We applied the Fisher exact test to categorical variables and the Student *t* test or the Mann-Whitney U test to continuous variables according to normality. We considered a 2-sided *P* value <.05 as significant.

We used a convenience sample and calculated the power post hoc. We defined 4 sensitivity analyses a priori according to whether the participant had prior knowledge of the NIHSS, had already followed a specific NIHSS course, had worked in either an intensive care unit or in a neurology or neurosurgery ward for more than 3 months, or had completed the learning path over multiple days. This was defined as more than 12 hours elapsed between initiation and completion of the course.

Finally, we performed univariate followed by multivariable linear regression to look for possible confounding factors.

Data Availability

Our curated data file is available on Mendeley Data [33].

Results

Out of 158 potential participants, 75 (47.5%) completed the trial (Figure 11). Table 1 details their characteristics.

After the first mailing (April 28, 2020), 21 students completed the trial. The first reminder (May 11, 2020) led 29 more students to complete the course, while another 25 participated after the second and last reminder (May 18, 2020).

Figure 11. Study flowchart.

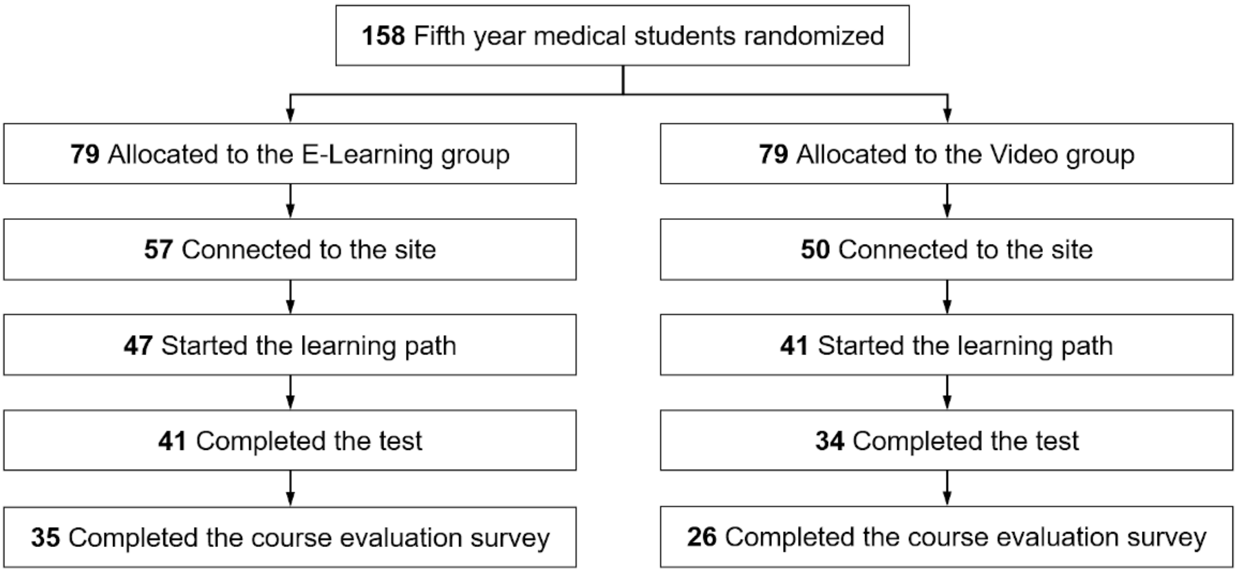


Table 1. Participant data (N=75).

Characteristic	Value	
	Video group (n=34)	E-Learning group (n=41)
Age, median (Q1-Q3)	24 (23-25)	24 (23-24)
Prior knowledge of NIHSS[a] application, n (%)	3 (9)	3 (7)
Specific NIHSS course followed, n (%)	4 (12)	2 (5)
E-learning NIHSS course followed, n (%)	0 (0)	0 (0)
Had worked in intensive care unit or neurology ward, n (%)	0 (0)	1 (2)
Missing data, n (%)	2 (6)	2 (5)

[a] NIHSS: National Institutes of Health Stroke Scale.

Participants who followed the e-learning module performed better than those who followed the video (38 correct answers, 95% CI 37-39, vs 35, 95% CI 34-36, *P*<.001). Participants in the e-learning group scored better on five elements than the video group: key NIHSS concepts (*P*=.02), the consciousness – global item (*P*<.001), the facial palsy item (*P*=.04), the ataxia item (*P*=.03) and the sensory item (*P*=.04). There was no such effect in the video group. Detailed results are shown in Table 2.

The rate of “very satisfied” participants was higher in the e-learning group (14/35, 40%; 95% CI 24%-56%) versus the video group (4/26, 15%; 95% CI 5%-25%, *P*=.02) (Figure 12).

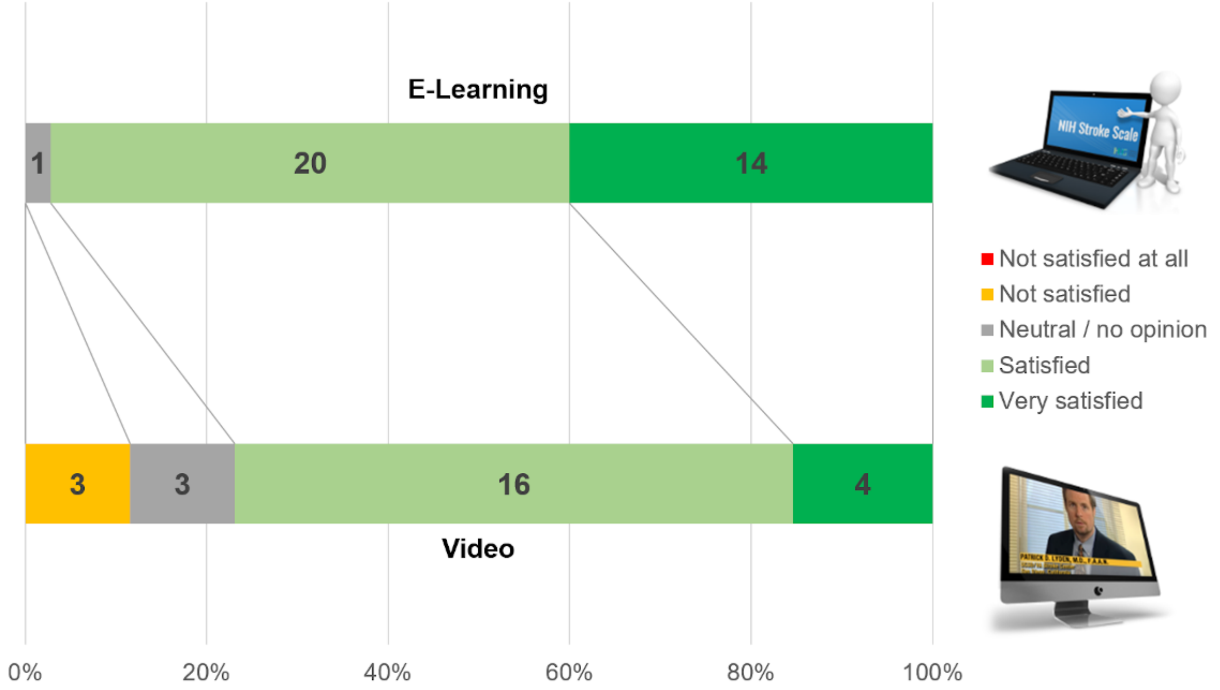
Table 2. Quiz results.

Item	Video group (n=34)	E-Learning group (n=41)	*P* value
Overall score, mean (SD)	35 (3)	38 (3)	<.001
Overall score, 95% CI	34-36	37-39	N/A[a]
Detailed results by item, median (Q1-Q3)			
Key NIHSS[b] concepts	5 (4-5)	5 (5-5)	.02
Consciousness – Global	2 (2-2)	3 (2-3)	<.001
Consciousness – Questions	3 (2-3)	3 (3-3)	.70
Consciousness – Commands	2 (2-3)	3 (2-3)	.06
Gaze	2 (2-3)	3 (2-3)	.34
Visual	2 (2-2)	2 (2-2)	.23
Facial Palsy	1 (0-2)	2 (1-2)	.04
Motor arm	4 (4-5)	5 (4-5)	.17
Motor leg	5 (4-6)	5 (4-5)	.23
Ataxia	1 (1-1)	1 (1-2)	.03
Sensory	3 (2-3)	3 (3-3)	.04
Language	1 (1-2)	1 (1-1)	.63
Dysarthria	2 (2-2)	2 (2-2)	.07
Extinction and inattention	2 (2-3)	2 (2-3)	.14

[a] N/A: not applicable.

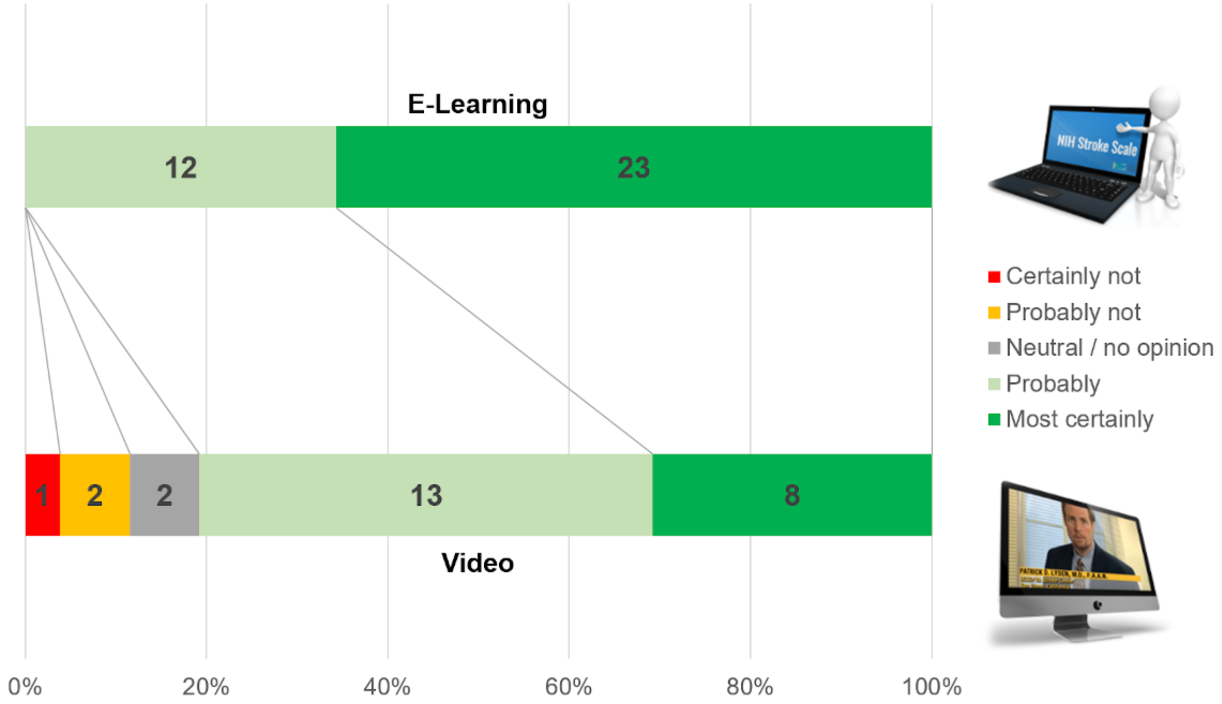
[b] NIHSS: National Institutes of Health Stroke Scale.

Figure 12. Satisfaction of the participants in the e-learning and video groups regarding the learning method.



Although the precise total learning time dedicated to either method could not be assessed due to the study design, no statistical difference regarding the perceived duration of the course was identified (80% (28/35) adequate in the e-learning group, 95% CI 67%-93%, vs 65% (17/26) 95% CI 47%-83%, *P*=.25). There was no significant difference regarding the perceived difficulty of the course, as 84% (16/19) (95% CI 68%-100%) found it “easy or very easy” in the e-learning group versus 53% (8/15) (95% CI 28%-78%) in the video group (*P*=.07). Participants who followed the e-learning method were more likely to recommend it to a colleague; 23/35 participants (66%) answered “Yes, most certainly” (95% CI 50%-82%), versus 8/26 (31%; 95% CI 13%-49%) in the video group, *P*=.007 (Figure 13). The proportions of participants in the two groups who followed the course over less than 12 hours were similar (58% (14/24) in the e-learning group, 95% CI 38%-78%, versus 52% (17/33) in the video group, 95% CI 35 to 69, *P*=.79).

Figure 13. Likelihood that the participants in the e-learning and video groups would recommend the course.



The post hoc calculation showed a power of 97%. None of the 4 preplanned sensitivity analyses showed any major changes in the direction of the effect. The multivariable linear regression only showed a minor change in the coefficient (<15%), confirming these results.

Discussion

Principal Findings

In this study, asynchronous distance learning using a highly interactive e-learning module yielded better results than following the traditional didactic video on the web. The superiority of a previous version of this module has already been established in Swiss paramedics following an onsite computer-based course [23]. The present study confirms the generalizability of these findings when using this method for asynchronous distance learning in a different population of learners, namely, fifth year medical students. Indeed, although paramedics follow a 3-year curriculum focusing on critical emergencies, baseline knowledge and understanding of neurosciences should be higher in students on the verge of obtaining their master of medicine degree [34]. This assumption is supported by the median score of the control group, which was higher by two points in this study than it was in paramedics [23].

The shortcomings we had identified in the previous iteration of the e-learning module seem to have been addressed, as embedding cutscenes from the original video into every chapter of the module improved the impact of the module on knowledge acquisition. The use of short videos associated with active learning activities such as guiding questions or interactive elements has been shown to enhance knowledge acquisition and retention [35]. Interactivity itself is also known to improve both engagement and performance in medical students [36,37].

Slightly less than half of all potential participants completed their allocated learning path. Considering that the learning material was optional and that students’ summative assessments of this semester were replaced by formative assessments, the participation rate is rather encouraging given the global lack of incentive. More encouraging still is the proportion of students who would recommend the course to their peers, as such mechanisms may increase students' involvement [38]. As many medical students actively helped on the front lines during the crisis, some of them may have been prevented from participating in this study owing to their high workload [39].

The quiz shown to the participants upon completion of the learning material included not only the full evaluation of 3 different stroke patients, but also 5 general questions we had designed Multimedia Appendix 2. While this could be considered as a potential bias in the study design, these questions were solely linked to key elements and basic principles of the NIHSS, and their understanding is essential to the correct application of the scale Multimedia Appendix 2. Our aim was indeed to evaluate whether knowledge acquisition was different when presenting similar content in different learning formats.

As the overall score regarding these questions was high in both groups, and as other significant results favored the e-learning method, there is little probability that these 5 initial questions induced a bias.

In many hospitals, the NIHSS is commonly used to triage stroke victims and help reduce both door-to-CT (CT: computed tomography) and door-to-needle times [9]. Decreasing these times is associated with better neurocognitive and functional outcomes [40]. Moreover, the adoption of a common scale between different specialists seems necessary to improve reproducibility and avoid the misinterpretation that can result from the use of different scores [41]. Swift acquisition and mastery of the NIHSS is therefore an essential skill for medical students, as most will be required to take care of stroke patients during their residency while working in the emergency department or in the neurology department. This is further strengthened by the fact that medical students often perceive neurology as the most difficult medical discipline, and the development of negative perceptions toward this specialty could lead to avoidance mechanisms when considering a career or treating a patient [42]. We might therefore assume that any kind of stroke-directed educational program could help raise awareness in non-neurologist physicians and thus increase the rate of correct treatment while decreasing door-to-needle time.

Limitations and Strengths

This study has limitations that must be acknowledged. The main limitation is that we only measured immediate knowledge acquisition; we were unable to assess knowledge retention due to the study design and the limited timeframe. As this latter parameter is critical to the clinical application of the NIHSS, further studies will be needed to assess whether the e-learning method improves retention and leads to more accurate application of the scale. Moreover, the precise time taken to complete either learning method was not evaluated in this study. While it can be argued that dedicating more time to learning given content should yield better results, studies have shown that engagement is the most important factor regarding knowledge acquisition [43]. Although time to learning material completion is an interesting outcome, we chose not to record these data for two main reasons: risk of unblinding and technical limitations. As the time required to watch the video is fixed unless the participants elect to use the video commands, and as most of them chose not to use this option in a previous study [23], we thought it better not to risk unnecessarily unblinding the data analyst. The technical aspect was linked to the web-based learning management system and to the mode of delivery of the teaching material. As access to the university premises was barred during the study period, participants followed the learning material from many different locations. Interruptions in the learning process may therefore have occurred; however, we had no means of recording recurrent short breaks in the group that followed the e-learning module, as pauses may also have resulted from taking notes, mulling over the content, or simply rereading some of the written paragraphs to better understand them. To mitigate this limitation, a sensitivity analysis comparing participants who completed the study path in less and more than 12 hours was performed. Reassuringly, no difference was noted.

Despite these limitations, this study also has several strengths, including the randomization, the blinding mechanisms, the electronic data acquisition, the originality of the learning method and its mode of delivery in the context of the COVID-19 pandemic.

Conclusion

Compared to the traditional didactic video, a highly interactive e-learning module enhances distant NIHSS knowledge acquisition in medical students.

Acknowledgments

The authors would like to thank Mr Daniel Scherly from the UGFM for his technical assistance as well as for dispatching the emails and reminders on our behalf.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was not registered. The authors explained that their study included no patients and no health outcomes were recorded. According to the ICMJE rules, if the purpose is to examine the effect only on the providers rather than patients, then registration is not necessary. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
Mailing template.
[PDF File (Adobe PDF File), 224 KB - jmir_v23i1e23594_app1.pdf]

Multimedia Appendix 2
Original questions used in the 50-question quiz.
[PDF File (Adobe PDF File), 276 KB - jmir_v23i1e23594_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2197 KB - jmir_v23i1e23594_app3.pdf]

References

1. Ahmed H, Allaf M, Elghazaly H. COVID-19 and medical education. Lancet Infect Dis 2020 Jul;20(7):777-778. [doi: 10.1016/S1473-3099(20)30226-7] [Medline: 32213335]
2. Hall AK, Nousiainen MT, Campisi P, Dagnone JD, Frank JR, Kroeker KI, et al. Training disrupted: Practical tips for supporting competency-based medical education during the COVID-19 pandemic. Med Teach 2020 May 25;42(7):756-761. [doi: 10.1080/0142159x.2020.1766669] [Medline: 32450049]
3. Chen CH, Mullen AJ. COVID-19 Can Catalyze the Modernization of Medical Education. JMIR Med Educ 2020 Jun 12;6(1):e19725 [FREE Full text] [doi: 10.2196/19725] [Medline: 32501809]
4. Naidoo N, Akhras A, Banerjee Y. Confronting the Challenges of Anatomy Education in a Competency-Based Medical Curriculum During Normal and Unprecedented Times (COVID-19 Pandemic): Pedagogical Framework Development and Implementation. JMIR Med Educ 2020 Oct 07;6(2):e21701 [FREE Full text] [doi: 10.2196/21701] [Medline: 32873536]
5. Klasen JM, Meienberg A, Nickel C, Bingisser R. SWAB team instead of SWAT team: Medical students as a frontline force during the COVID-19 pandemic. Med Educ 2020 Sep 17;54(9):860-860 [FREE Full text] [doi: 10.1111/medu.14224] [Medline: 32403176]
6. Harrington RA, Elkind MS, Benjamin IJ. Protecting Medical Trainees on the COVID-19 Frontlines Saves Us All. Circulation 2020 May 05;141(18):e775-e777. [doi: 10.1161/circulationaha.120.047454] [Medline: 32250654]
7. Olum R, Kajjimu J, Kanyike AM, Chekwech G, Wekha G, Nassozi DR, et al. Perspective of Medical Students on the COVID-19 Pandemic: Survey of Nine Medical Schools in Uganda. JMIR Public Health Surveill 2020 Jun 19;6(2):e19847 [FREE Full text] [doi: 10.2196/19847] [Medline: 32530815]
8. Flotte T, Larkin A, Fischer M, Chimienti S, DeMarco D, Fan P, et al. Accelerated Graduation and the Deployment of New Physicians During the COVID-19 Pandemic. Acad Med 2020 Oct 09;95(10):1492-1494 [FREE Full text] [doi: 10.1097/ACM.0000000000003540] [Medline: 32520751]
9. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, American Heart Association Stroke Council. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke 2018 Mar;49(3):e46-e110. [doi: 10.1161/STR.0000000000000158] [Medline: 29367334]
10. Montaner J, Barragán-Prieto A, Pérez-Sánchez S, Escudero-Martínez I, Moniche F, Sánchez-Miura JA, et al. Break in the Stroke Chain of Survival due to COVID-19. Stroke 2020 Aug;51(8):2307-2314. [doi: 10.1161/strokeaha.120.030106] [Medline: 32466738]
11. Mowla A, Doyle J, Lail N, Rajabzadeh-Oghaz H, Deline C, Shirani P, et al. Delays in door-to-needle time for acute ischemic stroke in the emergency department: A comprehensive stroke center experience. J Neurol Sci 2017 May 15;376:102-105. [doi: 10.1016/j.jns.2017.03.003] [Medline: 28431590]
12. Ross D, National Neuroscience Curriculum Initiative “Quarantine Curriculum” Committee. Creating a "Quarantine Curriculum" to Enhance Teaching and Learning During the COVID-19 Pandemic. Acad Med 2020 Aug;95(8):1125-1126 [FREE Full text] [doi: 10.1097/ACM.0000000000003424] [Medline: 32744816]
13. Ashokka B, Ong SY, Tay KH, Loh NHW, Gee CF, Samarasekera DD. Coordinated responses of academic medical centres to pandemics: Sustaining medical education during COVID-19. Med Teach 2020 Jul 13;42(7):762-771. [doi: 10.1080/0142159X.2020.1757634] [Medline: 32401085]
14. Iancu AM, Kemp MT, Alam HB. Unmuting Medical Students' Education: Utilizing Telemedicine During the COVID-19 Pandemic and Beyond. J Med Internet Res 2020 Jul 20;22(7):e19667 [FREE Full text] [doi: 10.2196/19667] [Medline: 32614776]
15. Qandil A, Abdel-Halim H. Distance e-Learning is Closer than Everybody Thought: A Pharmacy Education Perspective. Health Prof Educ 2020 Sep;6(3):301-303. [doi: 10.1016/j.hpe.2020.05.004]
16. Longhurst GJ, Stone DM, Dulohery K, Scully D, Campbell T, Smith CF. Strength, Weakness, Opportunity, Threat (SWOT) Analysis of the Adaptations to Anatomical Education in the United Kingdom and Republic of Ireland in Response to the Covid-19 Pandemic. Anat Sci Educ 2020 May;13(3):301-311 [FREE Full text] [doi: 10.1002/ase.1967] [Medline: 32306550]
17. Remtulla R. The Present and Future Applications of Technology in Adapting Medical Education Amidst the COVID-19 Pandemic. JMIR Med Educ 2020 Jul 17;6(2):e20190 [FREE Full text] [doi: 10.2196/20190] [Medline: 32634107]
18. McGowan BS. The Rise and Stall of eLearning: Best Practices for Technology-Supported Education. J Contin Educ Nurs 2015 Jul;46(7):292-294. [doi: 10.3928/00220124-20150619-11] [Medline: 26154667]
19. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. Med Teach 2008 Jun;30(5):455-473. [doi: 10.1080/01421590802108331] [Medline: 18576185]
20. Lee L, Chao Y, Huang C, Fang J, Wang S, Chuang C, et al. Cognitive Style and Mobile E-Learning in Emergent Otorhinolaryngology-Head and Neck Surgery Disorders for Millennial Undergraduate Medical Students: Randomized Controlled Trial. J Med Internet Res 2018 Feb 13;20(2):e56 [FREE Full text] [doi: 10.2196/jmir.8987] [Medline: 29439943]

21. Vaona A, Banzi R, Kwag KH, Rigon G, Cereda D, Pecoraro V, et al. E-learning for health professionals. Cochrane Database Syst Rev 2018 Jan 21;1:CD011736 [FREE Full text] [doi: 10.1002/14651858.CD011736.pub2] [Medline: 29355907]
22. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, et al. Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. Stroke 1994 Nov;25(11):2220-2226. [doi: 10.1161/01.str.25.11.2220] [Medline: 7974549]
23. Koka A, Suppan L, Cottet P, Carrera E, Stuby L, Suppan M. Teaching the National Institutes of Health Stroke Scale to Paramedics (E-Learning vs Video): Randomized Controlled Trial. J Med Internet Res 2020 Jun 09;22(6):e18358 [FREE Full text] [doi: 10.2196/18358] [Medline: 32299792]
24. Fouh E, Breakiron DA, Hamouda S, Farghally MF, Shaffer CA. Exploring students learning behavior with an interactive etextbook in computer science courses. Computers in Human Behavior 2014 Dec;41:478-485. [doi: 10.1016/j.chb.2014.09.061]
25. Kopp V, Stark R, Fischer M. Fostering diagnostic knowledge through computer-supported, case-based worked examples: effects of erroneous examples and feedback. Med Educ 2008 Aug;42(8):823-829. [doi: 10.1111/j.1365-2923.2008.03122.x] [Medline: 18564096]
26. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
27. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
28. 810.30 Federal Act on Research involving Human Beings (Human Research Act, HRA). The Federal Council of Switzerland. 2011 Sep 30. URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html [accessed 2020-05-12]
29. Suppan L, Suppan M. NIHSS EtuMed Study Platform. URL: https://nihss-etumed.cardiomobile.ch [accessed 2020-05-18]
30. Verschueren S, van Aalst J, Bangels A, Toelen J, Allegaert K, Buffel C, et al. Development of CliniPup, a Serious Game Aimed at Reducing Perioperative Anxiety and Pain in Children: Mixed Methods Study. JMIR Serious Games 2019 Jun 01;7(2):e12429 [FREE Full text] [doi: 10.2196/12429] [Medline: 31199333]
31. Suppan M, Gartner B, Golay E, Stuby L, White M, Cottet P, et al. Teaching Adequate Prehospital Use of Personal Protective Equipment During the COVID-19 Pandemic: Development of a Gamified e-Learning Module. JMIR Serious Games 2020 Jun 12;8(2):e20173 [FREE Full text] [doi: 10.2196/20173] [Medline: 32516115]
32. NIHSS E-Learning Modules. URL: https://www.nihss-study.ch [accessed 2020-08-16]
33. Suppan L, Suppan M, Stuby L, Carrera E, Savoldelli G. NIHSS - Medical Students Dataset. Mendeley Data. URL: https://data.mendeley.com/datasets/r3xcwph3rd/1 [accessed 2020-01-11]
34. Safdieh JE, Govindarajan R, Gelb DJ, Odia Y, Soni M. Core curriculum guidelines for a required clinical neurology experience. Neurology 2019 Feb 22;92(13):619-626. [doi: 10.1212/wnl.0000000000007187] [Medline: 30796141]
35. Brame CJ. Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. CBE Life Sci Educ 2016;15(4) [FREE Full text] [doi: 10.1187/cbe.16-03-0125] [Medline: 27789532]
36. Stevens NT, Holmes K, Grainger RJ, Connolly R, Prior A, Fitzpatrick F, et al. Can e-learning improve the performance of undergraduate medical students in Clinical Microbiology examinations? BMC Med Educ 2019 Nov 07;19(1):408 [FREE Full text] [doi: 10.1186/s12909-019-1843-0] [Medline: 31699068]
37. Wade SWT, Moscova M, Tedla N, Moses DA, Young N, Kyaw M, et al. Adaptive Tutorials Versus Web-Based Resources in Radiology: A Mixed Methods Analysis of Efficacy and Engagement in Senior Medical Students. Acad Radiol 2019 Oct;26(10):1421-1431. [doi: 10.1016/j.acra.2019.02.021] [Medline: 31047791]
38. Ang L, Breyer Y, Pitt J. Course recommendation as a construct in student evaluations: will students recommend your course? Stud High Educ 2016 Jul 05;43(6):944-959. [doi: 10.1080/03075079.2016.1199543]
39. Soled D, Goel S, Barry D, Erfani P, Joseph N, Kochis M, et al. Medical Student Mobilization During a Crisis: Lessons From a COVID-19 Medical Student Response Team. Acad Med 2020 Sep;95(9):1384-1387 [FREE Full text] [doi: 10.1097/ACM.0000000000003401] [Medline: 32282373]
40. Sauser K, Levine DA, Nickles AV, Reeves MJ. Hospital variation in thrombolysis times among patients with acute ischemic stroke: the contributions of door-to-imaging time and imaging-to-needle time. JAMA Neurol 2014 Sep 01;71(9):1155-1161. [doi: 10.1001/jamaneurol.2014.1528] [Medline: 25023407]
41. Purrucker JC, Hametner C, Engelbrecht A, Bruckner T, Popp E, Poli S. Comparison of stroke recognition and stroke severity scores for stroke detection in a single cohort. J Neurol Neurosurg Psychiatry 2015 Sep;86(9):1021-1028. [doi: 10.1136/jnnp-2014-309260] [Medline: 25466259]
42. Abushouk AI, Duc NM. Curing neurophobia in medical schools: evidence-based strategies. Med Educ Online 2016;21:32476 [FREE Full text] [doi: 10.3402/meo.v21.32476] [Medline: 27680578]
43. Deater-Deckard K, Chang M, Evans ME. Engagement states and learning from educational games. New Dir Child Adolesc Dev 2013 Mar 08;2013(139):21-30. [doi: 10.1002/cad.20028] [Medline: 23483690]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CSV: comma-separated values
CT: computed tomography
NIHSS: National Institutes of Health Stroke Scale
UGFM: University of Geneva Faculty of Medicine

Edited by R Kukafka; submitted 17.08.20; peer-reviewed by KH Miller; comments to author 09.09.20; revised version received 09.09.20; accepted 10.01.21; published 15.01.21.

Please cite as:
Suppan M, Stuby L, Carrera E, Cottet P, Koka A, Assal F, Savoldelli GL, Suppan L
Asynchronous Distance Learning of the National Institutes of Health Stroke Scale During the COVID-19 Pandemic (E-Learning vs Video): Randomized Controlled Trial
J Med Internet Res 2021;23(1):e23594
URL: http://www.jmir.org/2021/1/e23594/
doi:10.2196/23594
PMID:33428581

Original Paper

SARS-CoV-2 Surveillance in the Middle East and North Africa: Longitudinal Trend Analysis

Lori Post[1], PhD; Emily Marogi[2], BSc; Charles B Moss[3], PhD; Robert Leo Murphy[4], MD; Michael G Ison[5], MD, MSc; Chad J Achenbach[5], MD, MPH; Danielle Resnick[6], PhD; Lauren Singh[1], MPH; Janine White[1], MA; Michael J Boctor[2], BSc; Sarah B Welch[1], MPH; James Francis Oehmke[1], PhD

[1]Buehler Center for Health Policy and Economics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[3]Institute of Food and Agricultural Sciences, University of Florida, Gainsville, FL, United States
[4]Insitute of Global Health, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[5]Division of Infectious Disease, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[6]International Food Policy Research Institute, Washington DC, DC, United States

Corresponding Author:
Lori Post, PhD
Buehler Center for Health Policy and Economics
Feinberg School of Medicine
Northwestern University
420 E Superior
Chicago, IL, 60611
United States
Phone: 1 203 980 7107
Email: lori.post@northwestern.edu

Abstract

Background: The COVID-19 pandemic has disrupted the lives of millions and forced countries to devise public health policies to reduce the pace of transmission. In the Middle East and North Africa (MENA), falling oil prices, disparities in wealth and public health infrastructure, and large refugee populations have significantly increased the disease burden of COVID-19. In light of these exacerbating factors, public health surveillance is particularly necessary to help leaders understand and implement effective disease control policies to reduce SARS-CoV-2 persistence and transmission.

Objective: The goal of this study is to provide advanced surveillance metrics, in combination with traditional surveillance, for COVID-19 transmission that account for weekly shifts in the pandemic speed, acceleration, jerk, and persistence to better understand a country's risk for explosive growth and to better inform those who are managing the pandemic. Existing surveillance coupled with our dynamic metrics of transmission will inform health policy to control the COVID-19 pandemic until an effective vaccine is developed.

Methods: Using a longitudinal trend analysis study design, we extracted 30 days of COVID-19 data from public health registries. We used an empirical difference equation to measure the daily number of cases in MENA as a function of the prior number of cases, the level of testing, and weekly shift variables based on a dynamic panel data model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R.

Results: The regression Wald statistic was significant (χ^2_5=859.5, *P*<.001). The Sargan test was not significant, failing to reject the validity of overidentifying restrictions (χ^2_{294}=16, *P*=.99). Countries with the highest cumulative caseload of the novel coronavirus include Iran, Iraq, Saudi Arabia, and Israel with 530,380, 426,634, 342,202, and 303,109 cases, respectively. Many of the smaller countries in MENA have higher infection rates than those countries with the highest caseloads. Oman has 33.3 new infections per 100,000 population while Bahrain has 12.1, Libya has 14, and Lebanon has 14.6 per 100,000 people. In order of largest to smallest number of cumulative deaths since January 2020, Iran, Iraq, Egypt, and Saudi Arabia have 30,375, 10,254, 6120, and 5185, respectively. Israel, Bahrain, Lebanon, and Oman had the highest rates of COVID-19 persistence, which is the number of new infections statistically related to new infections in the prior week. Bahrain had positive speed, acceleration, and jerk, signaling the potential for explosive growth.

Conclusions: Static and dynamic public health surveillance metrics provide a more complete picture of pandemic progression across countries in MENA. Static measures capture data at a given point in time such as infection rates and death rates. By including speed, acceleration, jerk, and 7-day persistence, public health officials may design policies with an eye to the future. Iran, Iraq, Saudi Arabia, and Israel all demonstrated the highest rate of infections, acceleration, jerk, and 7-day persistence, prompting public health leaders to increase prevention efforts.

(J Med Internet Res 2021;23(1):e25830) doi:10.2196/25830

KEYWORDS

COVID-19; SARS-CoV-2 surveillance; wave two; second wave; global COVID-19 surveillance; MENA public health surveillance; MENA COVID-19; Middle East and North Africa surveillance metrics; dynamic panel data; MENA econometrics; MENA SARS-CoV-2; Middle East and North Africa COVID-19 surveillance system; MENA COVID-19 transmission speed; MENA COVID-19 transmission acceleration; COVID-19 transmission deceleration; COVID-19 transmission jerk; COVID-19 7-day lag; SARS-CoV-2; Arellano-Bond estimator; generalized method of moments; GMM; Bahrain; Iran; Iraq; Israel; Jordan; Kuwait; Lebanon; Oman; Qatar; Saudi Arabia; Syria; United Arab Emirates; Yemen; Algeria; Djibouti; Egypt; Libya; Morocco; Tunisia

Introduction

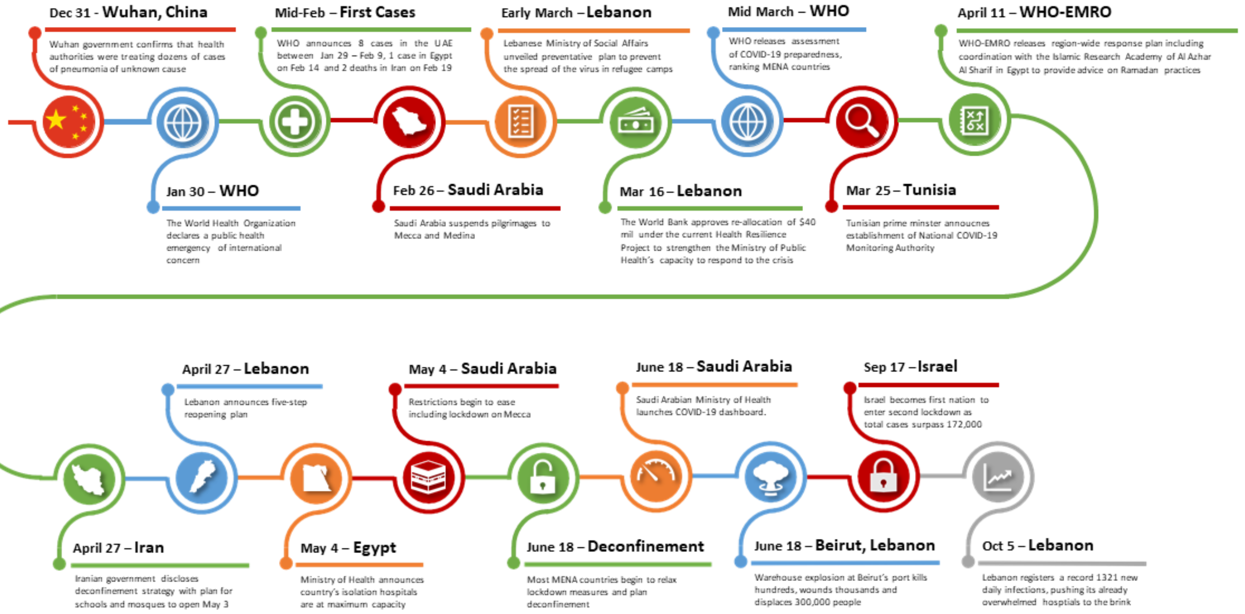
Background

SARS-CoV-2, the novel coronavirus that causes the disease COVID-19, first presented in December 2019 in Wuhan City, China, and was declared a public health emergency of international concern on January 30, 2020, as it spread quickly around the globe through human-to-human transmission [1-4]. According to the World Health Organization (WHO), the first reports of COVID-19 in the Middle East and North Africa (MENA) included 8 cases in the United Arab Emirates between January 29, 2020, and February 9, 2020 [5] (Figure 1). It has since spread to every country in the region [6]. MENA is divided by large wealth disparities and regional conflicts, increasing risks to the COVID-19 pandemic that has no political or religious agenda [6]. Despite varying levels of health system preparedness in MENA countries, early implementation of strict containment measures helped to limit the spread of SARS-CoV-2. As of October 28, the WHO reported 43,540,739 confirmed cases of COVID-19, including 1,160,650 deaths worldwide [7]. There have been 2,982,597 confirmed cases in MENA, resulting in 75,737 deaths [7].

Many countries use a combination of containment and mitigation strategies, including isolation of cases, contact tracing, social distancing, border closures, masking, hand and surface hygiene, and travel restrictions [8]. In the absence of a vaccine, worldwide systemic public health surveillance that can reliably track cases and identify where outbreaks will occur is needed to inform disease control policy for COVID-19 prevention [4,9]. Global SARS-CoV-2 surveillance necessitates dividing the globe into separately surveyed regions. The World Bank, an international financial institution that aims to reduce poverty and increase sustainable prosperity in nations throughout the globe, divides the world into regions based on shared geographic, cultural, and historical qualities. The MENA region comprises Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, the United Arab Emirates, Yemen, Algeria, Djibouti, Egypt, Libya, Morocco, and Tunisia [10].

Responses to COVID-19 in MENA have ranged from restrictive temporary lockdowns to denial and lack of organization [6,11]. Lessons learned from the severe acute respiratory syndrome (SARS), the Middle East respiratory syndrome (MERS), and the 2009 H1N1 outbreaks inform the current pandemic [12]. For example, in Saudi Arabia, the Ministry of Health established the Command and Control System and the Saudi Center for Disease Control and Prevention shortly after the rise of MERS [12]. Some countries repurposed surveillance resources to encourage COVID-19 containment. This is best illustrated in the United Arab Emirates where the "Oyoon" ("eyes" in Arabic) surveillance camera program, originally developed to track and prevent crime, was repurposed to check the temperatures of those passing by and to ensure people are social distancing [13].

Figure 1. Middle East and North Africa (MENA) timeline. WHO: World Health Organization; UAE: United Arab Emirates; EMRO: Regional Office for the Eastern Mediterranean.



Dec 31 - Wuhan, China
Wuhan government confirms that health authorities were treating dozens of cases of pneumonia of unknown cause

Jan 30 – WHO
The World Health Organization declares a public health emergency of international concern

Mid-Feb – First Cases
WHO announces 8 cases in the UAE between Jan 29 – Feb 9, 1 case in Egypt on Feb 14 and 2 deaths in Iran on Feb 19

Feb 26 – Saudi Arabia
Saudi Arabia suspends pilgrimages to Mecca and Medina

Early March – Lebanon
Lebanese Ministry of Social Affairs unveiled preventative plan to prevent the spread of the virus in refugee camps

Mar 16 – Lebanon
The World Bank approves re-allocation of $40 mil under the current Health Resilience Project to strengthen the Ministry of Public Health's capacity to respond to the crisis

Mid March – WHO
WHO releases assessment of COVID-19 preparedness, ranking MENA countries

Mar 25 – Tunisia
Tunisian prime minster annoucnes establishment of National COVID-19 Monitoring Authority

April 11 – WHO-EMRO
WHO-EMRO releases region-wide response plan including coordination with the Islamic Research Academy of Al Azhar Al Sharif in Egypt to provide advice on Ramadan practices

April 27 – Iran
Iranian government discloses deconfinement strategy with plan for schools and mosques to open May 3

April 27 – Lebanon
Lebanon announces five-step reopening plan

May 4 – Egypt
Ministry of Health announces country's isolation hospitals are at maximum capacity

May 4 – Saudi Arabia
Restrictions begin to ease including lockdown on Mecca

June 18 – Deconfinement
Most MENA countries begin to relax lockdown measures and plan deconfinement

June 18 – Saudi Arabia
Saudi Arabian Ministry of Health launches COVID-19 dashboard.

June 18 – Beirut, Lebanon
Warehouse explosion at Beirut's port kills hundreds, wounds thousands and displaces 300,000 people

Sep 17 – Israel
Israel becomes first nation to enter second lockdown as total cases surpass 172,000

Oct 5 – Lebanon
Lebanon registers a record 1321 new daily infections, pushing its already overwhelmed hosptials to the brink

Economics

Countries in MENA face a dual crisis from the COVID-19 pandemic and the collapse of oil prices [14]. These concurrent crises are exacerbated by structural economic challenges including large and inefficient public sectors, uncompetitive business environments, high youth and female unemployment, governance challenges, and regional conflicts [14]. Given these challenges, the World Bank released a report recommending a coordinated regional MENA trade integration framework that lays the foundation for global value chain integration and thus helps MENA's economic prospects [14]. Income disparities abound in MENA, which is home to the wealthy Persian Gulf nations and yet is one of the only regions in the world where extreme poverty has been rising since 2011, particularly in regions with violent conflicts [6,15]. Thus, as countries in MENA combat the COVID-19 pandemic, they do so in the context of falling oil prices and disparities in wealth and political stability.

Violent conflicts significantly weakened the health infrastructure in several countries across MENA, resulting in poor health worker capacity [12,16]. However, outside of conflict regions, MENA countries have some of the lowest proportion of health workers in the world with Egypt, Morocco, Iraq, Yemen, and Djibouti subject to low levels of doctors and nurses [11]. The health system capacity in MENA is facility based, which limits the surge capacity needed to respond to a crisis such as the COVID-19 pandemic [11]. Some MENA countries have limited public health infrastructure, including Egypt, Iraq, Djibouti, and Yemen that spend less than 5% of their government funding on health [11,17].

Middle Eastern populations have a high disease burden, which suggests a reduction in routine health service utilization results in increased mortality, a trend expected to hold true for COVID-19 [11,12]. Food insecurity poses a significant public health concern in countries with high refugee per capita concentration, such as Lebanon and Jordan [18-20]. Refugee populations are particularly vulnerable to COVID-19 as they are often concentrated in densely populated camps with poor hygiene measures and fragmented access to health care [12].

On August 4, 2020, a deadly warehouse explosion, caused by 2750 tons of ammonium nitrate stored at Beirut's port, killed hundreds, wounded and displaced thousands, left 5 hospitals in the area either nonfunctional or only partially functional, and destroyed 17 containers of WHO essential medical supplies [21-23]. The seeming carelessness with which this destructive material was stored for 6 years fueled public outrage and antigovernment protests throughout the city [24]. Since then, Lebanon has seen a 220% increase in COVID-19 cases [25]. As countries ease restrictions and a second wave of COVID-19 threatens to undo progress, it is particularly important to accurately track cases across MENA [26]. This is evident in countries such as Israel [27]. Ideally, advanced tracking and estimation of COVID-19 transmission will inform the implementation of public health policies. Though general public health surveillance is helpful and provides a proxy of the pandemic, surveillance data suffer from significant bias due to undercounts, reporting delays, testing errors, dearth of testing, asymptomatic carriers, and other types of data contamination [7,28-35]. In fact, surveillance systems are predicated on the fact that they tend to include the more severe cases and suffer from incomplete case ascertainment by missing the mild cases and undiagnosed infections and deaths.

Objective

The objective of our research is to provide additional surveillance metrics to add to the public health arsenal using dynamic panel modeling and method of moments to minimize sampling bias by measuring significant negative or positive weekly shifts in the increase, decrease, or plateaued transmission of SARS-CoV-2. We will apply novel indicators derived specifically to inform policy makers about the COVID-19 pandemic [3,36], including weekly shifts in the pandemic, speed, acceleration and deceleration, jerk, and 7-day persistence. These dynamic metrics will better inform public health policy when combined with standard surveillance measures.

Methods

This study relies on a longitudinal trend analysis of data collected from the Foundation for Innovative New Diagnostics (FIND) [37]. FIND compiles data from multiple sources across individual websites, statistical reports, and press releases; data for the most recent 8 weeks were accessed from the GitHub repository [38]. This resulted in a panel of 19 MENA countries with 30 days in each panel (n=570). An empirical difference equation was specified in which the number of new positive cases in each country on each day is a function of prior numbers of cases, level of testing, and weekly shift variables that measure whether the contagion was growing faster, at the same pace, or slower than the previous weeks. This resulted in a dynamic panel data model that was estimated using the generalized method of moments (GMM) approach by implementing the Arellano-Bond estimator in R (The R Foundation) [3,36,39].

Results

MENA Regression Results

Regression results are presented for 17 MENA countries in Table 1. Weekly surveillance data in Tables 2-7 are based on these regressions.

The regression Wald statistic was significant (χ^2_5=859.5, *P*<.001). The Sargan test was not significant, failing to reject the validity of overidentifying restrictions (χ^2_{294}=16, *P*=.99)

Table 1. Arellano-Bond dynamic panel data modeling of the number of daily infections reported by country, October 5-18, 2020.

Variable	Statistic	*P* value
L7Pos[a]	*r*=0.57	<.001
Cumulative tests	*r*=0.00012	<.001
Weekend	*r*=0.04	.95
Wald statistic for regression	χ^2_5= 859.5	<.001
Sargan statistic for validity	χ^2_{294}=16	.99

[a] L7Pos: the statistical impact of the 7-day lag of speed on today's value of speed. New cases per day tend to have an echo effect 7 days later, similar to the echo effect in the population pyramid caused by the baby boom. Reported as the weekly average number of new cases per day that are attributable to the weekly average of the 7-day lag of the number of new cases per day.

Tables 2 and 3 contain standard surveillance metrics for the weeks of October 5-11 and October 12-18. These metrics include new daily COVID-19 cases, cumulative COVID-19 cases, 7-day moving averages, rate of infection, new deaths, cumulative deaths, 7-day moving average of death rates, and rates of death per 1,000,000 population.

Table 2. Static surveillance metrics for the week of October 5-11, 2020.

Country	New COVID-19 cases, n	Cumulative COVID-19 cases, n	7-day moving average of new cases	Rate of infection	New deaths, n	Cumulative deaths, n	7-day moving average of death rate	Rate of death
Algeria	132	53,072	133.71	0.31	6	1801	5.86	0.01
Bahrain	327	75,614	421.71	19.92	2	275	2.14	0.12
Djibouti	0	5423	0.57	0.00	0	61	0.00	0.00
Egypt	129	104,516	119.00	0.13	12	6052	10.14	0.01
Iran	3822	500,075	4043.29	4.61	251	28,544	226.71	0.30
Iraq	2206	402,330	3312.71	5.61	62	9852	64.71	0.16
Israel	618	290,493	3388.29	6.83	39	1980	37.29	0.43
Jordan	928	24,926	1326.57	9.19	10	191	12.86	0.10
Lebanon	1010	53,568	1298.00	14.73	4	459	7.57	0.06
Libya	1026	42,712	843.29	15.14	8	631	5.57	0.12
Morocco	2563	152,404	2733.14	7.03	33	2605	39.29	0.09
Oman	1761	105,890	660.00	35.40	29	1038	8.71	0.58
Qatar	207	127,985	212.43	7.31	1	220	0.57	0.04
Saudi Arabia	323	339,267	411.43	0.94	25	5043	24.00	0.07
Tunisia	1297	32,556	1475.14	11.09	22	478	22.43	0.19
United Arab Emirates	1096	106,229	1061.14	11.22	2	445	2.71	0.02
Region	17,445	2,417,060	18,569.14	4.35	506	59,675	470.57	0.13

Table 3. Static surveillance metrics for the week of October 12-18, 2020.

Country	New weekly COVID-19 cases, n	Cumulative COVID-19 cases, n	7-day moving average of new cases	Rate of infection	New weekly deaths, n	Cumulative deaths, n	7-day moving average of death rate	Rate of deaths per 100k
Algeria	199	54,402	190	0.46	10	1856	7.86	0.02
Bahrain	331	77,902	326.86	20.17	7	300	3.57	0.43
Djibouti	7	5459	5.14	0.72	0	61	0.00	0.00
Egypt	127	105,424	129.71	0.13	11	6120	9.71	0.01
Iran	3890	530,380	4329.29	4.69	252	30,375	261.57	0.30
Iraq	3110	426,634	3472	7.91	56	10,254	57.43	0.14
Israel	339	303,109	1802.29	3.74	19	2209	32.71	0.21
Jordan	1520	37,573	1806.71	15.05	15	345	22.00	0.15
Lebanon	1002	62,286	1245.43	14.62	3	520	8.71	0.04
Libya	945	48,790	868.29	13.94	26	725	13.43	0.38
Morocco	2721	173,632	3032.57	7.46	50	2928	46.14	0.14
Oman	1657	109,953	580.43	33.31	30	1101	9.00	0.60
Qatar	204	129,431	206.57	7.20	1	224	0.57	0.04
Saudi Arabia	348	342,202	419.29	1.02	20	5185	20.29	0.06
Tunisia	0	40,542	1140.86	0.00	0	626	21.14	0.00
United Arab Emirates	1215	115,602	1339.00	12.44	4	463	2.57	0.04
Region	17,615	2,563,321	18,397.14	4.39	504	63,292	516.71	0.13

In Table 3, we see the countries with the highest cumulative caseload of the novel coronavirus, which includes Iran, Iraq, Saudi Arabia, and Israel with 530,380, 426,634, 342,202, and 303,109 cases, respectively. To eliminate the week-end effect or delayed reporting, we look at 7-day moving average to understand the 7-day average number of cases. In the second week of our study, October 12-18, Iran reported 4329, Israel reported 1802, Iraq reported 3472, and Morocco reported 3033 new cases per day (weekly averages). The next closest country in terms of average new cases only has a third of those cases. We find that many of the smaller countries in MENA have higher infection rates than those countries with the highest caseload. For example, Oman had 33.3 new infections per 100,000 population while Bahrain had 12.1, Libya had 14, and Lebanon had 14.6. In order of most to least number of cumulative deaths since January 2020, Iran, Iraq, Egypt, and Saudi Arabia had 30,375, 10,254, 6120, and 5185, respectively. These countries are outliers when comparing them to the rest of the region, which all have less than 3000 cumulative deaths

The novel surveillance metrics are presented in Tables 4 and 5 and Figure 2 [40] and Multimedia Appendices 1-5.

Table 4. Novel surveillance metrics for the week of October 5-11, 2020.

Country	Speed[a]	Acceleration[b]	Jerk[c]	7-day persistence effect on speed[d]
Algeria	0.31	0.00	0.00	0.20
Bahrain	25.70	–0.22	0.48	16.37
Djibouti	0.06	–0.01	0.00	0.08
Egypt	0.12	0.00	0.00	0.07
Iran	4.88	0.03	–0.03	2.49
Iraq	8.43	–0.36	0.12	6.15
Israel	37.43	–2.70	2.83	32.14
Jordan	13.13	0.05	–0.14	5.76
Lebanon	18.93	0.05	–0.09	9.77
Libya	12.44	0.64	0.75	5.34
Morocco	7.49	0.20	–0.10	3.48
Oman	13.27	–2.65	–2.65	6.25
Qatar	7.50	0.24	0.23	4.06
Saudi Arabia	1.20	–0.03	–0.02	0.76
Tunisia	12.61	0.01	–3.82	4.26
United Arab Emirates	10.86	0.08	0.23	6.11
Region	5.35	–0.09	–0.07	3.11

[a] Daily positives per 100k (weekly average of new daily cases per 100k).

[b] Day-to-day change in the number of positives per day, weekly average, per 100k.

[c] Week-over-week change in acceleration, per 100k.

[d] New cases per day per 100k attributed to new cases 7 days ago.

Table 5. Novel surveillance metrics for the week of October 12-18, 2020.

Country	Speed[a]	Acceleration[b]	Jerk[c]	7-day persistence effect on speed[d]
Algeria	0.44	0.02	0.00	0.18
Bahrain	19.92	0.03	1.61	14.64
Djibouti	0.53	0.10	0.06	0.03
Egypt	0.13	0.00	0.00	0.07
Iran	5.22	0.01	–0.03	2.78
Iraq	8.83	0.33	0.01	4.80
Israel	19.91	–0.44	1.36	21.32
Jordan	17.89	0.84	0.46	7.48
Lebanon	18.17	–0.02	0.44	10.79
Libya	12.81	–0.17	0.50	7.09
Morocco	8.31	0.06	–0.06	4.27
Oman	11.67	–0.30	–0.30	7.56
Qatar	7.29	–0.02	–0.30	4.27
Region	5.21	0.01	–0.06	3.05
Saudi Arabia	1.22	0.01	0.03	0.68
Tunisia	9.76	–1.58	–3.28	7.19
United Arab Emirates	13.70	0.17	–0.42	6.19
Region	5.21	0.01	–0.06	3.05

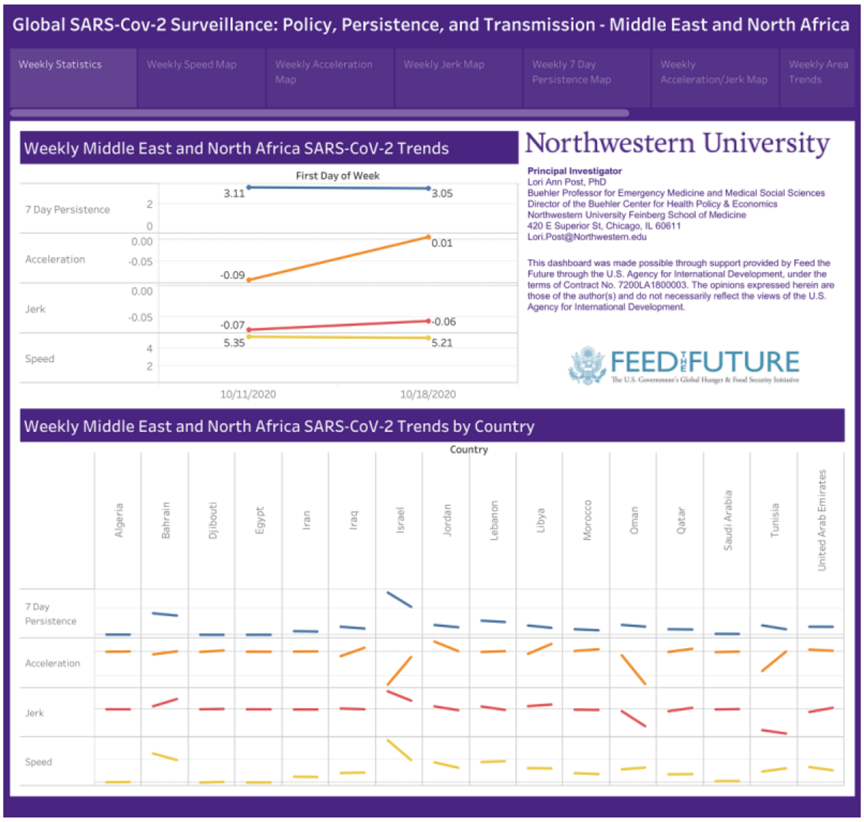
[a] Daily positives per 100k (weekly average of new daily cases per 100k).

[b] Day-to-day change in the number of positives per day, weekly average, per 100k.

[c] Week-over-week change in acceleration, per 100k.

[d] New cases per day per 100k attributed to new cases 7 days ago.

Figure 2. Middle East and North Africa weekly trends [40].



Principal Investigator
Lori Ann Post, PhD
Buehler Professor for Emergency Medicine and Medical Social Sciences
Director of the Buehler Center for Health Policy & Economics
Northwestern University Feinberg School of Medicine
420 E Superior St, Chicago, IL 60611
Lori.Post@Northwestern.edu

This dashboard was made possible through support provided by Feed the Future through the U.S. Agency for International Development, under the terms of Contract No. 7200LA1800003. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the U.S. Agency for International Development.

Overall, the MENA region was more stable than other global regions, but metrics between October 5-18 indicate pending growth. The speed of new infections decelerated during the week of October 5-11 and remained stable during the week of October 11-18. In addition, in the latter week, there was a slight negative jerk. The persistence rate slightly decreased from 3.11 to 3.05 per 100,000 population over the study period, which directly measures those new cases that are statistically related to the number of new infections 7 days earlier. While the surveillance metrics for the MENA region as a whole are promising, they are averages and thus we must look to those countries with increasing rates of speed, acceleration, jerk, and persistence to understand which countries have outbreaks and less control of the pandemic. In Table 5, we note that Bahrain, Israel, and Lebanon decreased in speed or remained level between October 5 and 18. Jordan and the United Arab Emirates had significant increases in rates of infection. Tunisia had the best shift in the pandemic that resulted in a decrease in speed week over week. In addition, the rates reversed course and decelerated from October 5-11 to October 12-18, and Tunisia experienced a negative jerk in infection rates 2 weeks in a row, indicating a strong downward trend in COVID-19 infections. For Israel, the 7-day persistence effect for the week of October 12-18 (21 new cases) slightly overpredicts the speed or the actual number of new cases (n=20); coupled with a declining persistence rate relative to the prior week of October 5-11, this is indicative of a rapid descent from a recent period of explosive growth. On the other end of the spectrum, Bahrain had positive speed, acceleration, and jerk, indicating the outbreak is increasing and is leading toward a worsening pandemic.

Table 6. Seven-day persistence difference.

Country	7-day persistence	
	October 5-11, 2020	October 12-18, 2020
Israel	32.14	21.32
Bahrain	16.37	14.64
Lebanon	9.77	10.79
Oman	6.25	7.56
Iraq	6.15	—[a]
Jordan	—	7.48

[a] Not applicable.

Israel, Bahrain, Lebanon, and Oman had the highest rate of COVID-19 persistence, which is the number of new infections statistically related to new infections from 7 days ago. These 3 countries had higher persistence rates 2 weeks in a row. Finally, Egypt has the highest population in MENA (Table 7) and the potential for growth but has remained stable across the 2 weeks of this study, indicating that COVID-19 is not accelerating.

Table 7. Most populous countries in the Middle East and North Africa.

Country	Population as of 2020, n
Egypt	102,334,404
Iran	83,992,949
Algeria	43,851,044
Iraq	40,222,493
Morocco	36,910,560

Discussion

Principal Findings

Analysis at the country level indicates there are some nations that should increase their public health efforts to gain control of the COVID-19 pandemic. Iran, Israel, Iraq, and Morocco had the highest reported 7-day average per 100,000 population, which is significantly higher than other nations in the region. Iran, Iraq, Saudi Arabia, and Israel had the highest caseload at the end of October 18. Looking toward the future, Jordan, Iraq, and the United Arab Emirates have the fastest acceleration in new COVID-19 infections while Bahrain and Israel have the largest upwards jerk in infections, which can lead to explosive growth. Iran began the pandemic with explosive growth but during the last week of this study ending on October 18, Iran's acceleration rate has leveled off and its jerk has reversed lower; however, given the number of new cases and population size, the country could easily flare up in new outbreaks, especially considering the second wave of COVID-19 infections has just begun.

Limitation

Data are limited by granularity and collection method. Data were collected at the country level, which precludes local analysis of surveillance trends. Moreover, data collection mechanisms differ by country and may even differ by region within a given country. These different methods lead to week-end effects, missing data points, and other contamination.

Comparison With Prior Work

This study is part of a broader research program at Northwestern University Feinberg School of Medicine, *The Global SARS-CoV-2 Surveillance Project: Policy, Persistence, & Transmission*. This research program developed novel surveillance metrics to include rates of speed, acceleration, jerk, and 7-day persistence [3,36]. We have also derived surveillance metrics for all global regions.

Conclusion

Static and dynamic public health surveillance tools provide a more complete picture of pandemic progression across countries and regions. While static measures capture data at a given point in time, like infection rates and death rates, they are less successful at assessing population health over a period of weeks or months. By including speed, acceleration, jerk, and 7-day persistence, public health officials may design policies with an eye to the future.

MENA countries with the highest risk all shared a number of characteristics according to the surveillance data. There was a definite positive shift between October 5-11 and October 12-18. Iran, Iraq, Saudi Arabia, and Israel all demonstrated the highest numbers of cumulative infections, acceleration, jerk, and 7-day persistence rates. Looking ahead, policy makers in these countries and the region at large should be concerned about growth in the already substantial number of cases over the short term. Given the substantial 7-day persistence rates of Israel, Bahrain, and Lebanon, it is imperative that efforts be made to target super spreader events. Analysis of subsequent surveillance data using both static and dynamic tools can help confirm the efficaciousness of new policies.

Acknowledgments

Partial support for this publication was provided by Feed the Future through the US Agency for International Development, under the terms of contract #7200LA1800003. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the US Agency for International Development.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Weekly MENA country statistics.
[PNG File , 329 KB - jmir_v23i1e25830_app1.png]

Multimedia Appendix 2
Weekly MENA SARS-CoV-2 trends.
[PNG File , 139 KB - jmir_v23i1e25830_app2.png]

Multimedia Appendix 3
Weekly MENA acceleration and jerk map.
[PNG File , 291 KB - jmir_v23i1e25830_app3.png]

Multimedia Appendix 4
Weekly MENA 7-day persistence map.
[PNG File , 265 KB - jmir_v23i1e25830_app4.png]

Multimedia Appendix 5
Weekly MENA speed map.
[PNG File , 258 KB - jmir_v23i1e25830_app5.png]

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
2. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
3. Oehmke J, Oehmke T, Singh L, Post L. Dynamic Panel Estimate-Based Health Surveillance of SARS-CoV-2 Infection Rates to Inform Public Health Policy: Model Development and Validation. J Med Internet Res 2020 Sep 22;22(9):e20924 [FREE Full text] [doi: 10.2196/20924] [Medline: 32915762]

4. Heymann DL, Shindo N. COVID-19: what is next for public health? The Lancet 2020 Feb;395(10224):542-545. [doi: 10.1016/s0140-6736(20)30374-3]
5. Update on COVID-19 in the Eastern Mediterranean Region. WHO EMRO. 2020 Feb 16. URL: http://www.emro.who.int/media/news/update-on-covid-19-in-the-eastern-mediterranean-region.html [accessed 2020-10-26]
6. Karamouzian M, Madani N. COVID-19 response in the Middle East and north Africa: challenges and paths forward. The Lancet Global Health 2020 Jul;8(7):e886-e887. [doi: 10.1016/s2214-109x(20)30233-3]
7. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-08-26]
8. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, et al. COVID-19: towards controlling of a pandemic. The Lancet 2020 Mar;395(10229):1015-1018. [doi: 10.1016/s0140-6736(20)30673-5]
9. Pubilc health surveillance. World Health Organization. 2014. URL: https://www.who.int/immunization/monitoring_surveillance/burden/vpd/en/ [accessed 2020-08-03]
10. Middle East and North Africa. The World Bank. 2020. URL: https://www.worldbank.org/en/region/mena [accessed 2020-08-03]
11. Duran D, Menon R. Mitigating the Impact of COVID-19 and Strengthening Health Systems in the Middle East and North Africa. World Bank. 2020. URL: https://openknowledge.worldbank.org/handle/10986/34238 [accessed 2020-12-14]
12. Sawaya T, Ballouz T, Zaraket H, Rizk N. Coronavirus Disease (COVID-19) in the Middle East: A Call for a Unified Response. Front Public Health 2020 May 19;8:209 [FREE Full text] [doi: 10.3389/fpubh.2020.00209] [Medline: 32574298]
13. Gambrell J. Virus projects renew questions about UAEs mass surveillance. The Associated Press. 2020. URL: https://apnews.com/article/e70e359d46391672a7a88f254cf9025b [accessed 2020-10-22]
14. Arezki R, Blanca; YF, Rachel; G, Romeo; N, Ha; N, Minh CM, et al. Middle East and North Africa Economic Update: Trading Together? Reviving Middle East and North Africa Regional Integration in the Post-COVID Era. The World Bank. 2020. URL: https://www.worldbank.org/en/region/mena/publication/mena-economic-update-trading-together-reviving-middle-east-and-north-africa-regional-integration-in-the-post-covid-era [accessed 2020-10-20]
15. Ncube M, Anyanwu JC, Hausken K. Inequality, Economic Growth and Poverty in the Middle East and North Africa (MENA). African Development Review 2014 Sep 16;26(3):435-453. [doi: 10.1111/1467-8268.12103]
16. Arezki R, Nguyen H. 4 Novel coronavirus hurts the Middle East and North Africa through many channels. Economics in the Time of COVID-19 2020:53-58 [FREE Full text]
17. Mehrara M, Fazaeli AA, Fazaeli AR. International Journal of Business Management and Economic Research 2012;3(1):425-428 [FREE Full text]
18. Jomaa L, Naja F, Cheaib R, Hwalla N. Household food insecurity is associated with a higher burden of obesity and risk of dietary inadequacies among mothers in Beirut, Lebanon. BMC Public Health 2017 Jun 12;17(1):567 [FREE Full text] [doi: 10.1186/s12889-017-4317-5] [Medline: 28606120]
19. Ghattas H, Sassine AJ, Seyfert K, Nord M, Sahyoun NR. Prevalence and Correlates of Food Insecurity among Palestinian Refugees in Lebanon: Data from a Household Survey. PLoS One 2015 Jun 22;10(6):e0130724 [FREE Full text] [doi: 10.1371/journal.pone.0130724] [Medline: 26098108]
20. Jomaa L, Naja F, Kharroubi S, Hwalla N. Prevalence and correlates of food insecurity among Lebanese households with children aged 4–18 years: findings from a national cross-sectional study. Public Health Nutr 2018 Dec 04;22(2):202-211. [doi: 10.1017/s1368980018003245]
21. Beirut explosion: Lebanon's government 'to resign' as death toll rises. BBC News. 2020 Aug 10. URL: https://www.bbc.com/news/world-middle-east-53720383 [accessed 2014-12-10]
22. Saldivia G, Sherlock R, Homsi N. Lebanon's Prime Minister Resigns, Citing An 'Earthquake' Of Upheaval After Explosion. NPR. 2020 Aug 10. URL: https://www.npr.org/2020/08/10/900878730/lebanons-prime-minister-resigns-citing-an-earthquake-of-upheaval [accessed 2020-08-10]
23. WHO requires US$ 15 million to support immediate Beirut blast response and ensure vital health and disease prevention activities are maintained. WHO EMRO. 2020 Aug 6. URL: https://bit.ly/2XowGHO [accessed 2020-08-26]
24. Beirut explosion: Angry protesters storm government ministries. BBC News. 2020 Aug 8. URL: https://www.bbc.com/news/world-middle-east-53704998 [accessed 2020-08-10]
25. Vohra A. Coronavirus Is Pushing Lebanon Over the Brink. Foreign Policy. 2020 Oct 8. URL: https://foreignpolicy.com/2020/10/08/coronavirus-is-pushing-lebanon-over-the-brink/ [accessed 2020-10-22]
26. OECD Policy Responses to Coronavirus (COVID-19): COVID-19 crisis response in MENA countries. OECD. 2020 Nov 6. URL: https://www.oecd.org/coronavirus/policy-responses/covid-19-crisis-response-in-mena-countries-4b366396/ [accessed 2020-08-26]
27. Teibel A. Israel to Enter Second Virus Lockdown After Bungled Reopen. Bloomberg. 2020 Sep 13. URL: https://www.bloomberg.com/news/articles/2020-09-13/israel-to-enter-second-virus-lockdown-after-bungled-reopen?utm_campaign=pol&utm_medium=bd&utm_source=applenews [accessed 2020-10-22]

28. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
29. Pettersson H, Manley B, Hernandez S. Tracking coronavirus' global spread. CNN. 2020. URL: https://www.cnn.com/interactive/2020/health/coronavirus-maps-and-cases/ [accessed 2020-08-26]
30. CoronaTracker. 2020. URL: https://www.coronatracker.com/about [accessed 2020-08-26]
31. The Visual and Data Journalism Team. Coronavirus Pandemic: Tracking the Global Outbreak. BBC News. 2020. URL: https://www.bbc.com/news/world-51235105 [accessed 2020-08-26]
32. Coronavirus Tracked: The Latest Figure as Countries Fight Covid-19 Resurgence. Financial Times. 2020. URL: https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938 [accessed 2020-08-26]
33. COVID-19 statistical report – Jordan. Ministry of Health: The Hashemite Kingdom of Jordan. 2020. URL: https://corona.moh.gov.jo/en [accessed 2020-08-26]
34. Middle Eastern Mediterranean Regional Office COVID-19 Dashboard. WHO Regional Office for the Eastern Mediterranean. 2020. URL: https://bit.ly/3q383g49 [accessed 2020-08-26]
35. Saudi Arabia COVID-19 dashboard. Ministry of Health–Kingdom of Saudi Arabia. 2020. URL: https://covid19.moh.gov.sa [accessed 2020-08-26]
36. Oehmke J, Moss C, Singh L, Oehmke T, Post L. Dynamic Panel Surveillance of COVID-19 Transmission in the United States to Inform Health Policy: Observational Statistical Study. J Med Internet Res 2020 Oct 5;22(10):e21955. [doi: 10.2196/21955]
37. SARS-CoV-2 test tracker. FIND. 2020. URL: https://www.finddx.org/covid-19/test-tracker/ [accessed 2020-10-23]
38. FINDCov19TrackerData. GitHub. 2020. URL: https://github.com/dsbbfinddx/FINDCov19TrackerData [accessed 2020-10-23]
39. Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 1982 Jul;50(4):1029. [doi: 10.2307/1912775]
40. Global SARS-CoV-2 Surveillance Policy, Persistence, and Transmission - Middle East and North Africa. Tableau Public. 2020 Oct 25. URL: https://public.tableau.com/profile/lori.post#!/vizhome/GlobalSARS-CoV-2SurveillancePolicyPersistenceandTransmission-MiddleEastandNorthAfrica/Story [accessed 2020-12-14]

Abbreviations

FIND: Foundation for Innovative New Diagnostics
GMM: generalized method of moments
MENA: Middle East and North Africa
MERS: Middle East respiratory syndrome
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 17.11.20; peer-reviewed by P Banik; comments to author 08.12.20; revised version received 08.12.20; accepted 10.12.20; published 15.01.21.

Please cite as:
Post L, Marogi E, Moss CB, Murphy RL, Ison MG, Achenbach CJ, Resnick D, Singh L, White J, Boctor MJ, Welch SB, Oehmke JF
SARS-CoV-2 Surveillance in the Middle East and North Africa: Longitudinal Trend Analysis
J Med Internet Res 2021;23(1):e25830
URL: http://www.jmir.org/2021/1/e25830/
doi:10.2196/25830
PMID:33302252

Original Paper

Unfolding the Determinants of COVID-19 Vaccine Acceptance in China

Fulian Yin[1], PhD; Zhaoliang Wu[1], MA; Xinyu Xia[1], BE; Meiqi Ji[1], BE; Yanyan Wang[1], ME; Zhiwen Hu[2], PhD

[1]Communication University of China, Beijing, China

[2]Zhejiang Gongshang University, Hangzhou, China

Corresponding Author:
Zhiwen Hu, PhD
Zhejiang Gongshang University
No 18 Xuezheng Street
Higher Education Zone
Hangzhou
China
Phone: 86 13676458899
Email: huzhiwen@zjgsu.edu.cn

Abstract

Background: China is at the forefront of global efforts to develop COVID-19 vaccines and has five fast-tracked candidates at the final-stage, large-scale human clinical trials testing phase. Vaccine-promoting policymaking for public engagement is a prerequisite for social mobilization. However, making an informed and judicious choice is a dilemma for the Chinese government in the vaccine promotion context.

Objective: In this study, public opinions in China were analyzed via dialogues on Chinese social media, based on which Chinese netizens' views on COVID-19 vaccines and vaccination were investigated. We also aimed to develop strategies for promoting vaccination programs in China based on an in-depth understanding of the challenges in risk communication and social mobilization.

Methods: We proposed a novel behavioral dynamics model, SRS/I (susceptible-reading-susceptible/immune), to analyze opinion transmission paradigms on Chinese social media. Coupled with a meta-analysis and natural language processing techniques, the emotion polarity of individual opinions was examined in their given context.

Results: We collected more than 1.75 million Weibo messages about COVID-19 vaccines from January to October 2020. According to the public opinion reproduction ratio (R_0), the dynamic propagation of those messages can be classified into three periods: the ferment period (R_{01}=1.1360), the revolution period (R_{02}=2.8278), and the transmission period (R_{03}=3.0729). Topics on COVID-19 vaccine acceptance in China include price and side effects. From September to October, Weibo users claimed that the vaccine was overpriced, making up 18.3% (n=899) of messages; 38.1% (n=81,909) of relevant topics on Weibo received likes. On the contrary, the number of messages that considered the vaccine to be reasonably priced was twice as high but received fewer likes, accounting for 25.0% (n=53,693). In addition, we obtained 441 (47.7%) positive and 295 (31.9%) negative Weibo messages about side effects. Interestingly, inactivated vaccines instigated more heated discussions than any other vaccine type. The discussions, forwards, comments, and likes associated with topics related to inactivated vaccines accounted for 53% (n=588), 42% (n=3072), 56% (n=3671), and 49% (n=17,940), respectively, of the total activity associated with the five types of vaccines in China.

Conclusions: Most Chinese netizens believe that the vaccine is less expensive than previously thought, while some claim they cannot afford it for their entire family. The findings demonstrate that Chinese individuals are inclined to be positive about side effects over time and are proud of China's involvement with vaccine development. Nevertheless, they have a collective misunderstanding about inactivated vaccines, insisting that inactivated vaccines are safer than other vaccines. Reflecting on netizens' collective responses, the unfolding determinants of COVID-19 vaccine acceptance provide illuminating benchmarks for vaccine-promoting policies.

(J Med Internet Res 2021;23(1):e26089) doi:10.2196/26089

KEYWORDS

COVID-19 vaccines; COVID-19 vaccination; affordability; efficacy; risk communication; evidence communication; social media; COVID-19; vaccine; communication; risk; acceptance; China; opinion; strategy; promotion

Introduction

Background

Vaccines have been proven to be an extremely effective means of dealing with epidemics in the past [1]. However, over the past decades, the antivaccine or antivaccination movement has taken root in Europe and the United States [2,3]. The antivaccine movement, which encourages vaccine hesitancy, has emerged as a significant public health problem, topping the list of threats to global health [4]. For example, it fueled the contagious measles outbreak of 2019 [5,6]. In addition, antivaccination misinformation spreads more quickly than positive counterparts [7]. Immediately after declaring COVID-19 as a pandemic, numerous conspiracy theories were shared through social media [8-10]. In Pakistan, for example, two well-known political figures expressed anti–COVID-19 vaccine sentiments to the local community and further encouraged vaccine hesitancy [11]. Neil et al [12] proposed a heuristic map of COVID-19 vaccines' online contentions, which revealed a multisided landscape of unprecedented intricacies about vaccines. The reasons for vaccine refusal are complex and vary by geographical and sociocultural contexts.

Many studies have shown that even vaccinated individuals may have substantial doubts and concerns regarding vaccination [13-15]. Many experts believe that vaccination programs are threatened by growing concerns in the population regarding vaccines' safety and efficacy [16-18]. According to previous estimates, less than 5%-10% of individuals have strong antivaccination convictions [19]. However, a more significant proportion could be categorized as being vaccine hesitant [20]. Vaccine hesitancy, defined as individual-level reluctance to receive vaccines, may be fueled by a spectrum of held views regarding vaccination spanning from cautious acceptors to outright deniers [21-23]. Amin et al [24] proposed that values-based messages appeal to core morality, influencing individuals' attitudes on topics like vaccination. They showed via two correlational studies that harm and fairness are not significantly associated with vaccine hesitancy, but purity and liberty are. In addition, politics and public trust may affect public perceptions of vaccine risks. Larson [25] discusses the following aspects: some risks of vaccines, such as side effects, provoke anxiety, reluctance, and rejection of vaccination; when vaccines are regulated, and sometimes mandated, by the government, it is resisted by those who feel their freedom is being imposed upon. Those who do not trust the government sometimes extend their distrust to vaccines produced by pharmaceutical companies, which will generate profits and incite public concern about the motives of vaccine production. By examining the antivaccine situation in Texas, United States, Martin [26] also concluded that the antivaccine community, at large, believes that vaccines are a tool used for government control that makes big pharmaceutical companies wealthy and have side effects that can cause lasting damage. Among the barriers to universal vaccination, misinformation regarding the benefits, medicinal composition, and adverse effects of vaccination limits individuals' understanding and overall buy-in [27]. Vaccine safety concerns continue to be an essential driver of decreasing vaccine uptake in most contexts [28-30]. Several reports indicate that people's opinions of vaccination have a significant influence on a vaccine's development and marketing.

COVID-19 has been demonstrated to have high human-to-human transmissibility [31-33]. The ability of SARS-CoV-2 to infect people through asymptomatic carriers is difficult to detect, making the disease a confounding public health challenge [34-36]. Therefore, vaccine development studies have been carried out by the research teams of various companies and universities worldwide [37,38]. Among them, Chinese research on COVID-19 vaccines is a special case, which covers almost all types of vaccines. However, COVID-19 vaccine development has incited heated discussions [39]. Different countries have varying attitudes toward vaccine development. For example, the Japanese government is considering free COVID-19 vaccination for all residents when it becomes available [26]. In California, United States, some individuals carried placards with antivaccine slogans at rallies to protest against the lockdown. Subsequently, antivaccine movements have also taken place in London, United Kingdom, and other cities [40,41]. In China, although more than 90,000 families have been affected by the epidemic [42], no such movement has taken place. Yet, it does not mean that there is no contention about COVID-19 vaccines and vaccination. Hence, it is of great importance to uncover Chinese people's collective propensities in social dialogues and to aid authorities in making reasonable and informed decisions.

There will be considerable variation by country in terms of COVID-19 vaccine acceptance [43]. To this end, this country-specific study aims to explore the paradigm of public engagement about COVID-19 vaccination to develop practical strategies of preparedness in order to mitigate the pandemic in China. This study investigates the trending topics related to COVID-19 vaccines on Weibo, and obtains public opinions and propensities related to COVID-19 vaccines, such as vaccine price and side effects.

Study Objectives

We aim to examine what Chinese netizens are concerned about in terms of COVID-19 vaccines and vaccination by profiling pertinent topics on the microblogging platform Weibo. We took random samples of more than 10 million Weibo messages from January to October 2020 to address the following research issues: the affordability of the COVID-19 vaccine candidates; the efficacy of the COVID-19 vaccine candidates; and propensities concerning COVID-19 vaccination. We also aimed to unveil the underlying motives behind these public appeals and explore potential strategies of preparedness for health and risk communication.

Methods

Data Collection

Weibo is thought of as a natural experiment that profiles social responses to Chinese public health preparedness. As the leading Chinese social media characterized by heterogeneous communities, it is a crucial public opinion platform in China. As of December 2018, Weibo had 462 million active users per month, which has increased by more than 70 million for 3 consecutive years, and had 200 million active users with 130 million words posted per day [44]. Chinese netizens regard this platform as a preferred outlet for expressing their demands and appeals [45]. More and more messages were posted, read, forwarded, and commented on than any other platform. Clusters of messages can be found on different topics marked by the hashtag symbol (#), which groups similar content. By organizing the same information into a topic, users can quickly find what they want to understand or express, thus resulting in large-scale participation. This mechanism is also a routine way of compiling comprehensive reflections of peoples' opinions.

In this study, we retrieved more than 1.75 million Weibo messages with approximately 21.17 billion links posted worldwide from January to October 2020. In addition, we classified the reliability of the messages being circulated. The messages were in 108 languages from around the world, but because of our data filtering and enrichment procedures, the largest fraction of analyzed messages point to Chinese-language sources (Simplified Chinese and Traditional Chinese). Additionally, for each message, verification was performed by Sina Corporation to clearly identify accounts of public interest and certify their authenticity, according to China's real-name verification policy for the use of microblogs. The findings reported in this study mainly captured the social behaviors of the Chinese-speaking portion of Weibo users, including domestic Chinese and those living abroad.

We utilized natural language processing to screen all Weibo topics about COVID-19 vaccines from the end of January to the beginning of October and obtained 989 topics. Of those, the typical, pertinent, and clustered topics are highlighted in Table 1.

Table 1. Topics related to the COVID-19 vaccine, with metadata on topic name, reading quantity, and date.

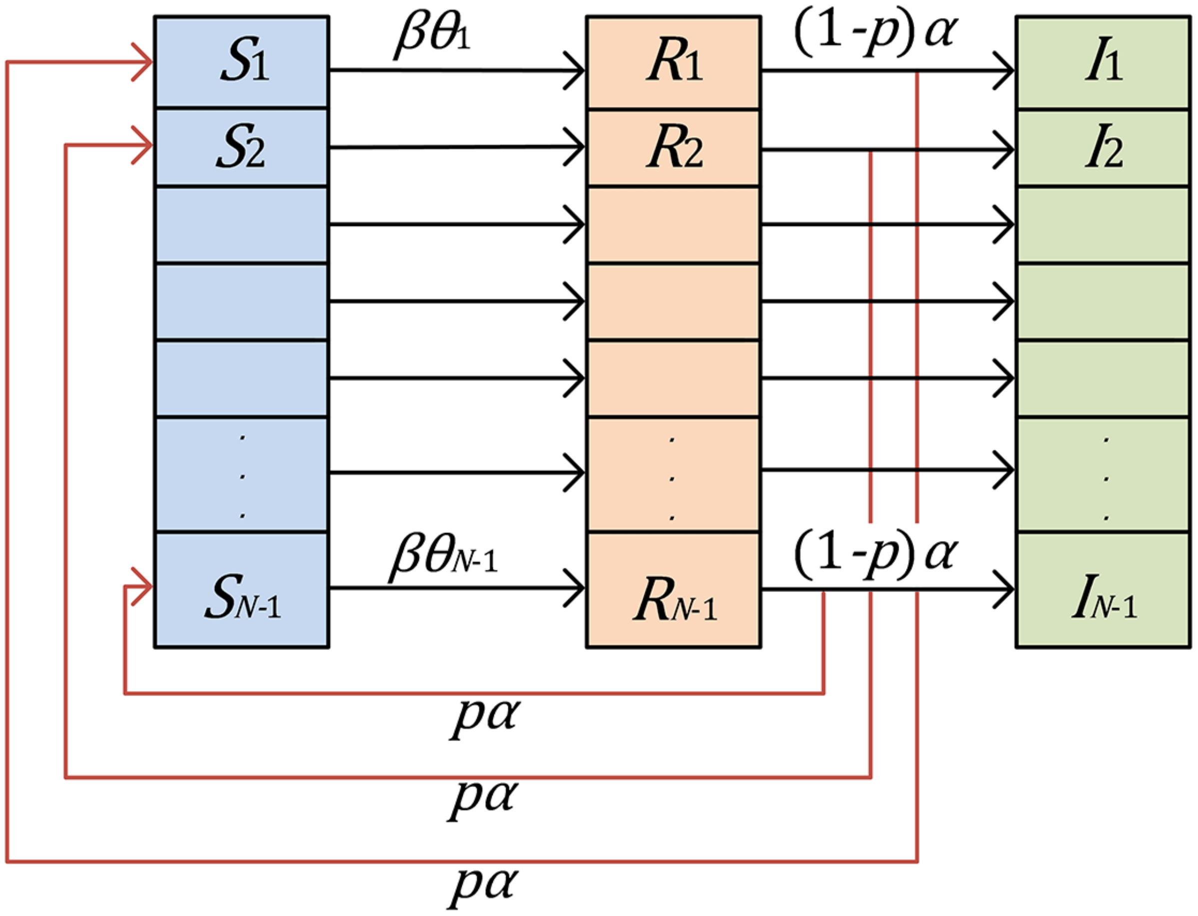
Topic name	Readers (million), n	Date
#COVID-19 vaccine could be available in early 2021#	820	February 9
#Wei Chen’s Team conducts the Phase I clinical trial of COVID-19 vaccine#	243.8	March 20
#When will the COVID-19 vaccine be available#	50.9	April 14
#China’s COVID-19 vaccine has entered Phase II clinical trial#	33	April 14
#The first participant of the COVID-19 vaccine has not yet shown adverse reactions#	21	April 14
#The world’s first COVID-19 inactivated vaccine#	41.2	April 19
#What is the COVID-19 inactivated vaccine#	25.2	April 21
#Chinese first COVID-19 inactivated vaccine entered Phase II clinical trial#	4566.4	April 24
#Phase I clinical trial of the first Chinese COVID-19 vaccine has good results#	934.5	May 22
#The safety and effectiveness of the COVID-19 inactivated vaccine have been verified#	76.9	May 29
#More than 2000 people received the COVID-19 inactivated vaccine injection#	479.6	May 30
#COVID-19 inactivated vaccine is expected to be available at the end of this year or early next year#	162.1	May 30
#China developed another kind of COVID-19 inactivated vaccine#	2000	June 9
#World’s first COVID-19 inactivated vaccine participant produces antibodies#	39.4	June 17
#Chinese COVID-19 vaccine will be launched as early as 2021#	37.9	June 18
#CNBG’s COVID-19 inactivated vaccine is not affected by virus mutation#	1276.2	June 19
#Chinese three COVID-19 vaccines complete Phase II clinical trials#	18,000	June 20
#Domestic COVID-19 inactivated vaccine launches international clinical phase Ⅲ trial#	84.2	June 23
#COVID-19 inactivated vaccine production workshop is completed in Wuhan#	7388.4	July 2
#WHO requires the protection period of the COVID-19 vaccine to be at least six months#	56.8	July 3
#Chinese COVID-19 vaccine Phase 2 clinical trial achieves good results#	131	July 21
#COVID-19 vaccine may be available at the end of the year#	345.4	July 22
#World first officially releases Phase II clinical data of COVID-19#	238.3	July 23
#The price of the COVID-19 vaccine will not exceed $40#	888.9	July 29
#Russian COVID-19 vaccine will be free of charge#	20,000	August 1
#The first COVID-19 inactivated vaccine workshop passed safety inspection#	44.9	August 5
#Beijing COVID-19 inactivated vaccine production workshop can be put into production at any time#	45.5	August 5
#Half of the COVID-19 vaccines in Phase III clinical trials come from China#	150.6	August 7
#Gates required the COVID-19 vaccine to be priced below $3#	963.7	August 8
#Research Institute refutes rumors of COVID-19 vaccine 498 yuan an injection#	127.4	August 13
#COVID-19 vaccine not yet be available#	3	August 13
#COVID-19 inactivated vaccine two injections less than one thousand yuan#	3081.6	August 18
#Domestic COVID-19 inactivated vaccine is expected to be available at the end of December#	131.6	August 18
#Russian second COVID-19 vaccine starts Phase 2 clinical trial#	70.1	August 18
#COVID-19 vaccine two injections 1000 yuan is too expensive#	24.2	August 19
#The price of COVID-19 vaccine can only be based on cost#	17,000	August 23
#How to price the COVID-19 vaccine#	254.2	August 23
#CNBG declared COVID-19 vaccine is likely to be available at the end of this year#	354.6	August 23
#National Health Commission claimed that the price of COVID-19 vaccine was lower than two injections of 1,000 yuan#	65.9	August 23
#Domestic COVID-19 inactivated vaccine first appears#	19,000	September 5
#Domestic COVID-19 inactivated vaccine appears in CIFTIS#	5766.4	September 6

Topic name	Readers (million), n	Date
#The COVID-19 vaccine will be priced based on factors such as consumers’ ability to pay#	134.5	September 7
#Domestic COVID-19 inactivated vaccine inoculates hundreds of thousands of people with zero infection#	2102.4	September 11
#Oxford vaccine volunteers have side effects#	769.1	September 9
#The COVID-19 vaccine produced by Pfizer in America has side effects#	17,000	September 16
#The COVID-19 inactivated vaccine is expected to be available at the end of the year#	4319.2	September 8
#The COVID-19 vaccine price in China will be within the scope of public acceptance#	11,000	September 2
#The COVID-19 inactivated vaccine is only one kilometer away from success#	23,000	September 2
#Two injections 600 yuan for domestic COVID-19 vaccine #	7737	September 3
#The price of domestic COVID-19 vaccine is released#	4413.4	September 2
#Four COVID-19 vaccines in China enter Phase III clinical trials#	428.6	September 2
#China formally joins COVAX#	23,000	October 9

Behavioral Dynamics Model

On social media, the propagation and inline influence of various topics are involved. To track the derivative development of topics related to COVID-19 vaccines in China, we propose the dynamics model SRS/I (susceptible-reading-susceptible / immune) based on a complex network to investigate the landscape of public opinion transmission (Figure 1) [46]. The SRS/I model promises to profile the collective propensities of different populations in terms of different topics across various times on social media.

Figure 1. Inspired by infectious disease models, a network model diagram of information dissemination is proposed to simulate the transmission of information among the susceptible state (S), the reading state (R), and the immunized state (I).



We considered a reading population of Weibo users for topics about the vaccine, stratified in terms of three distinct stages: the susceptible state (S), in which users are unaware of but susceptible to the information about the event; the reading state

(R), in which users have actively read information to influence other users; and the immunized state (I), in which users have read the information and can trigger a new round of reading activities on the same topic. We obtain the following SRS/I reading dynamics model of vaccines topics:

where β refers to the average exposure rate of a susceptible user who can read topics about the vaccine. Since an active reading user will contact an average number of βN users per unit time and the probability of a contacted user being a susceptible user is $S(t)/N$, the number of new reading users is $\beta\langle k\rangle N(S(t)/NR(t)) = \beta\langle k\rangle S(t)R(t)$. Users can become inactive to the same topic with an average inactive rate α, with $1/\alpha$ being the average duration where an R-user remains active in reading. The average number of inactive users will be $\alpha R(t)$ per unit time, among which $p\alpha R(t)$ will re-enter the susceptible state where exposure to another Weibo message within the same topic can initiate a new round of reading, and $(1-p)\alpha R(t)$ will proceed to the immunized state directly, in which p reflect the re-entering probability for a reading user who can trigger a new round of reading activities on the same topic. θ is a parameter related to the topics' dissemination network of topics pertaining to vaccines. In this paper, we only considered the average degree $\langle k\rangle$.

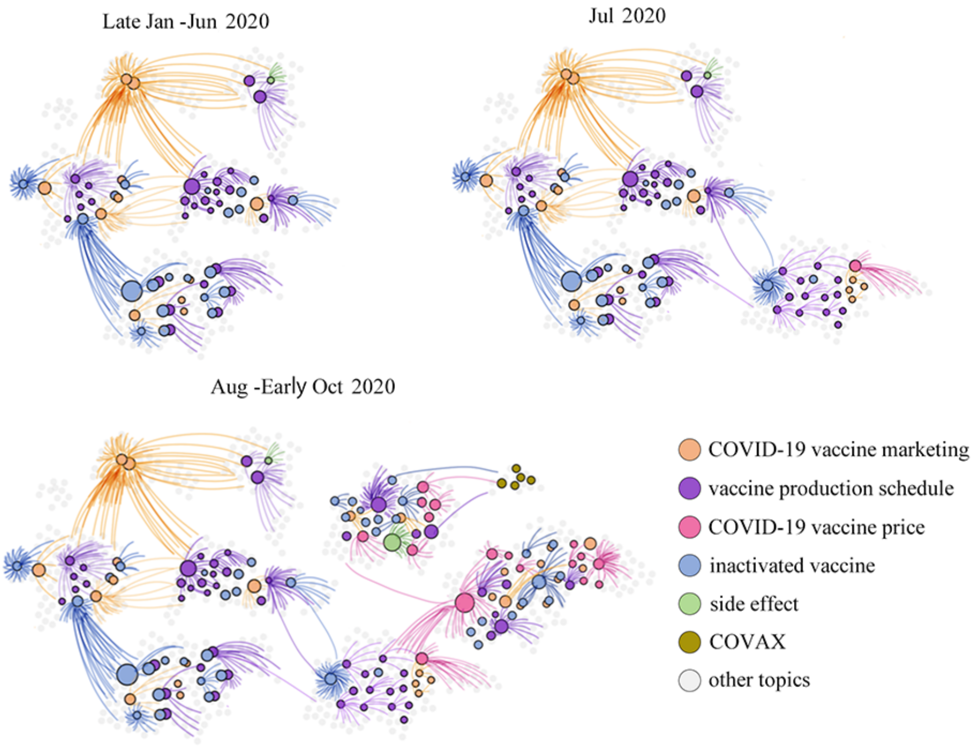
We extended the basic R_0 [47] of epidemiology to the field of information transmission. In our SRS/I dynamics model, we defined the public opinion R_0 as a measure of the potential impacts of topics in the initial propagation stage, which is given by $R'(0) = (\beta\langle k\rangle S_0 - \alpha)R(0)$, and the outbreak will never grow since $R'(0) = (\beta\langle k\rangle S_0 - \alpha)R(0) < 0$ due to the decrease of S. Then we deduce:

To further explore Chinese opinion about COVID-19 vaccines, we collected metadata, which includes the names and reading quantity of related topics from the end of January to the beginning of October, through an application programming interface provided by Weibo. Table 1 shows several specific topics related to COVID-19 vaccines with their posted date and reading quantity. The date can aid in identifying the continuity of topics over time, while the reading quantity could reflect the collective interest of the population.

The Landscape of Public Opinion Transmission

After numerical fitting and calculation, according to the basic R_0, we found that a paradigm shift emerged in July 2020. The R_0 value becomes large at this time, implying that Chinese netizens had become more interested in vaccines. Therefore, we divided the transmission process of vaccine-related topics into three periods: ferment period (stage 1, the end of January to June), evolution period (stage 2, July), and transmission period (stage 3, August to the beginning of October) (Figure 2). Their public opinion reproduction ratio is specifically expressed as R_{01}, R_{02}, and R_{03}, where R_{01}=1.1360, R_{02}=2.8278, and R_{03}=3.0729. It is evident that R_{03} is larger than R_{01} and R_{02} This is consistent with the fact that the topics in stage 3 have been disseminated more widely than those in stages 1 and 2. In stage 3, two sensitive topics came to our attention: vaccines' price and vaccines' side effects. Topics about vaccines' side effects (labels 44 and 45 in the Multimedia Appendix 1) reached more than 17,000,000,000 readers while topics about vaccine pricing (labels 24 and 25 in the Multimedia Appendix 1) have emerged in an endless stream and have been widely read. So, they incite continuous attention from Chinese netizens.

Figure 2. The landscape of public opinion transmission on vaccine-related topics. The colored dots represent the seven main topics. The lines between them represent mutual influence, and the density of the lines represents the degree of influence.



As shown in Figure 2, we obtained the metadata of 989 topics on Weibo. We considered the duration and content of topics that Chinese people are interested in. Topics were divided into seven categories based on content. We found topics connected with the same specific event are inline as time goes by. Namely, one topic, whether it has a strong positive or negative sentiment or is just a general announcement, may affect the sensitivity of Chinese netizens to the COVID-19 vaccine, thereby involving the population to facilitate the creation of new topics or a derivation of old ones.

Price Acceptance

We collected all messages and their likes on all price-related topics mentioned above. We set keywords and used the following formula to calculate the price tendency. After tokenization and the word extraction process, we obtain M words, which are synonyms for “expensive,” expressed as (E_1, E_2,…,E_M). By querying in the dictionary (Multimedia Appendix 2), we denoted the weight of the i_{th} word as a_i. Hence, the score of “expensive” is computed as:

Similarly, we obtained N words, synonyms for “cheap,” expressed as (C_1, C_2,…,C_N). After querying in the dictionary, the weight of the i_{th} word a_i was obtained. The score of “cheap” is expressed as:

The final expression of the price tendency score can be calculated as:

When $S>0$, the population regards the price as high, and when $S<0$, the collective attitudes toward price are acceptable.

Sentiment Polarity of Side Effects

We analyze the sentiment of 925 messages on topics related to side effects based on the general Chinese lexicon HowNet [48]. For the text sequence $x=\{w_1,\dots\dots,w_k,\dots\dots,w_K\}$, w_k indicates that the k-th word in sequence x, and K is the total word number for the sequence x. Then we obtained the corresponding sentimental values $Sen_x=\{S_U(w_1),\dots\dots, S_U(w_k),\dots\dots, S_U(w_K)\}$ of each word in x by sentiment lexicon rescores to synthesize the final sentence sentiment value S_x:

where $S_U(w_k)$ is the sentiment value for the k-th word in sequence x calculated by the sentiment lexicon we used, and S_x is the sentiment value of the sequence x. We turned S_x into polarity T_x as follows to easily judge the performance of our sentiment classification task:

where we count the sum of T_x of each case, denoted by NP, NN, and ZN, respectively. We further calculated the proportion of each sentiment tendency in all corpora:

where N is the total number of corpora, $pos(\%)$ and $neg(\%)$ are represented as positive and negative sentiment proportions, respectively.

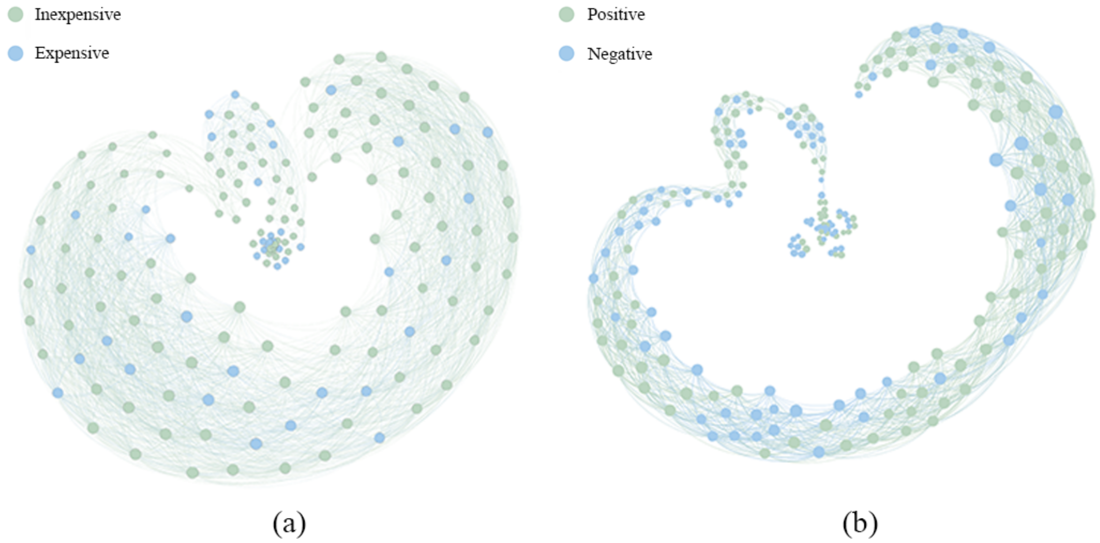
In the sampling inspection, due to the Chinese language's diversity, we validated the results of the sentiment lexicon–based method. We strictly followed the requirements of the double-blind experiment and invited three groups (A, B, and C) who had been trained to classify these Weibo messages by emotion. When the emotion judged by groups A and B was consistent, we took it as the correct result. When the judgment between groups A and B was inconsistent, group C made the final judgment. We marked positive Weibo messages as 1, and negative ones as –1, using the equations 10 and 11 to calculate manual labeling.

Results

Weibo's Attitude Influence Map

In the case of affordability and efficacy, we chose two topics—#The COVID-19 vaccine price can only be based on cost# (n=169 messages) and #The COVID-19 vaccine produced by Pfizer in America has side effects# (n=220 messages). We utilized the SRS/I model to draw the messages' attitude influence map (Figure 3). The colors represent the different attitudes. A topic is composed of individual Weibo messages, which have a content connection and time-sequence relationship. At first, the first Weibo message had the ability to promote subsequent publishing and dissemination. In Figure 3, we used points to represent Weibo messages, which appear clockwise in chronological order. In terms of price (Figure 3A), most people thought the pricing of vaccines was inexpensive. Nevertheless, it is worth noting that many still feel the pricing to be expensive, as seen by Weibo messages posted at various time periods. In terms of side effects (Figure 3B), positive Weibo messages dominated initially, but over time, positive and negative views alternated in being the majority, and the two views clashed during this time. Overall, positive dominates.

Figure 3. Weibo's attitude contagion map: (A) price and (B) side effects. Points represent Weibo messages that appear in the figure in clockwise order, ending with the point at the center of the figure.

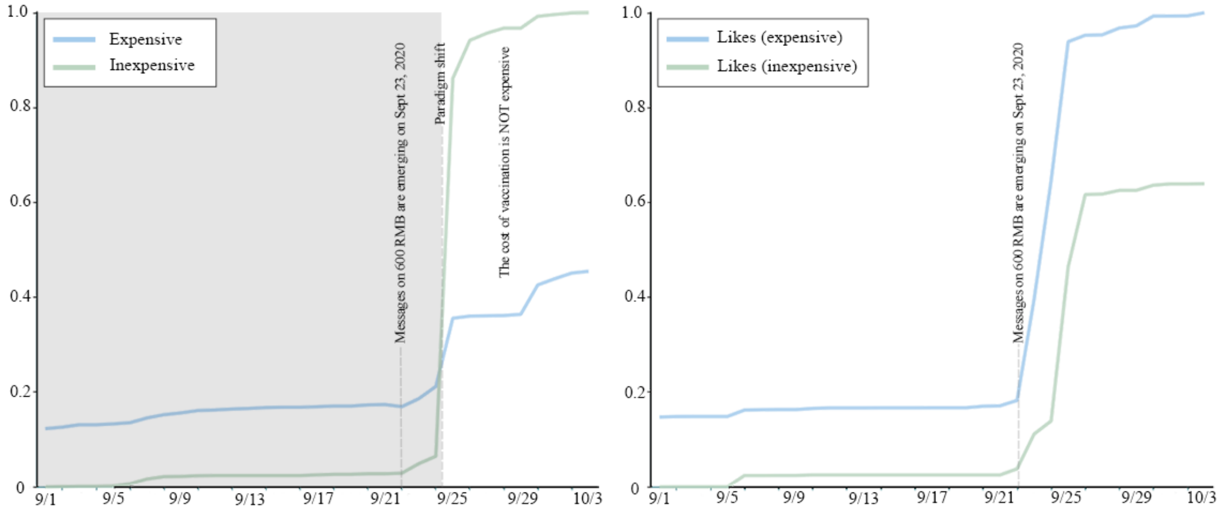


Affordability: The Price of Vaccines

We selected metadata during the period when the paradigm of price tendency dramatically changed and obtained 4925 related messages on Weibo. The final result is normalized (Figure 4).

Figure 4A shows that the collective attitudes toward COVID-19 vaccine pricing suddenly changed after September 23 from expensive to inexpensive. However, the population's likes displayed the opposite tendency, as shown in Figure 4B.

Figure 4. The paradigm shift in price tendency.



(a) (b)

On Weibo, the topic of COVID-19 vaccine pricing emerged in July 2020 and attracted hundreds of millions of onlookers in China. However, the number of messages that specifically address pricing is not very large.

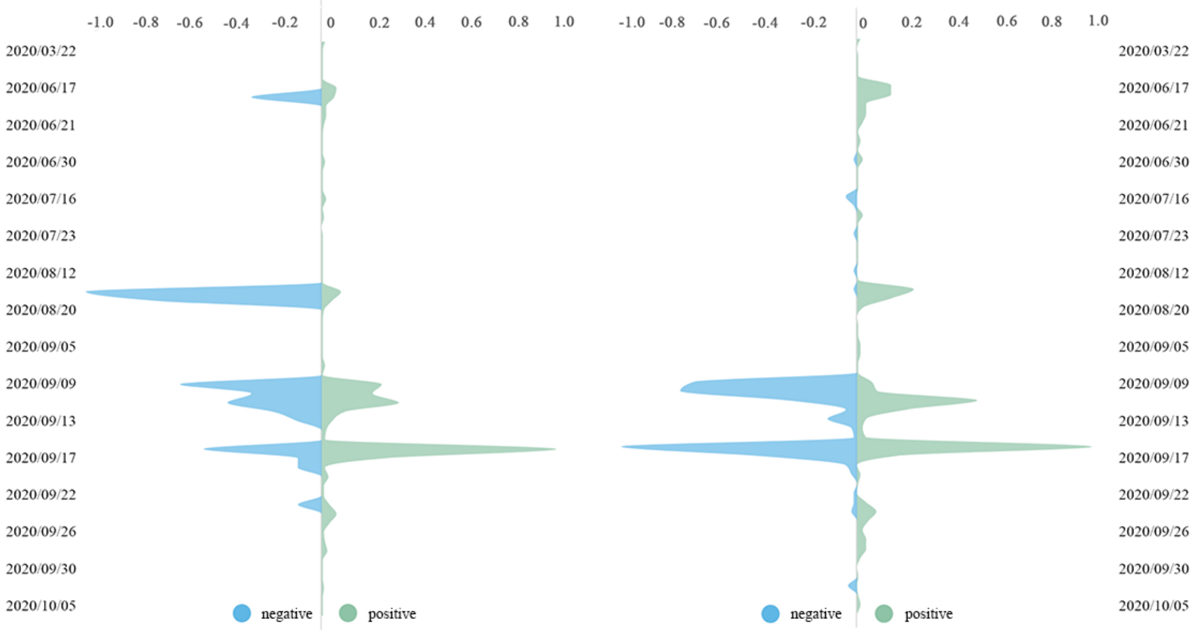
In August, the Chinese government claimed that the Chinese vaccine price would not exceed 1000 RMB (US $154). Therefore, the expected price was about 1000 RMB, and most people thought it was too expensive. Due to the one-child policy, young people may need to pay for their extended family. Although they believe that the cost of production is high, the cost of vaccination is not a small expenditure for ordinary families. After the Chinese COVID-19 vaccine fee was announced on September 23—600 RMB (US $90) for 2 shots—the number of relevant messages began to increase significantly (label 49 in Multimedia Appendix 1). As shown in Figure 4A, it is worth noting that the public had their own speculations about the prices of Chinese vaccines before this announcement. The public's attitudes reversed after this tipping point, and most people found the official price to be acceptable. Most people clicked "like" or posted Weibo messages to endorse previous views on Weibo rather than posting similar messages themselves. As shown in Figure 4B, the collective emotion polarity did not reverse with the increase in likes.

Certainly, the official price is still too expensive for some netizens. Comparatively speaking, citizens in developed countries have their own universal health insurance, or some countries promise to bear the costs of vaccinations. As a case in point, the Japanese and Russian governments have promised to pay to vaccinate its citizens (labels 25 and 52 in the Multimedia Appendix 1, respectively). Therefore, some Chinese netizens naturally hope that China will follow suit. The patterns in Figure 3A also confirm the above findings.

Efficacy: Side Effects of Vaccines

We plotted the lexicon-based sentiment classification results and the manually labeled sentiment classification method in Figure 5. Generally speaking, Chinese people are optimistic about vaccines. Interestingly, Chinese netizens seem to be accustomed to expressing their positive opinions in diverse ways. These utterances may undermine the accuracy rate of lexicon-based sentiment analysis, especially for negative sentiments (Figure 5). Potential uncertainties may cause a miscalculation of the accuracy rate of emotion-cause pair extraction and identification of ironic contexts. For example, comparing Figure 5A with Figure 5B, Chinese netizens tend to express their understanding of vaccine side effects using irony. In one Weibo post, for instance, a user wrote, "I think most people cannot even understand the title," while another wrote, "if you have a little common sense in pharmacology… Is it weird to have an adverse event (AE)? How can it be on the hot search?" These Weibo messages' authors believe that it is common sense for drugs to have side effects. They have a positive attitude toward side effects but scoff at those who have opposing opinions. However, these Weibo messages were judged neutral by the machine. There were also some Weibo messages judged to be negative, although their attitude toward side effects is positive. For example, one message read, "Even taking vitamins will increase the liver's metabolic burden. All drugs are somewhat toxic. If a drug has no side effects, it must be fake." They know the potential danger of side effects but think it is within the acceptable range. Thus, their attitude toward vaccines' side effects is positive. Due to irony, the results of the lexicon-based sentiment classification (Figure 5A) are inaccurate. Therefore, we only used the results obtained by manual labeling to determine Chinese netizens' attitudes toward side effects.

Figure 5. The collective attitudes toward side effects: (A) the results of the lexicon-based method and (B) the final results coupled with manual labeling.

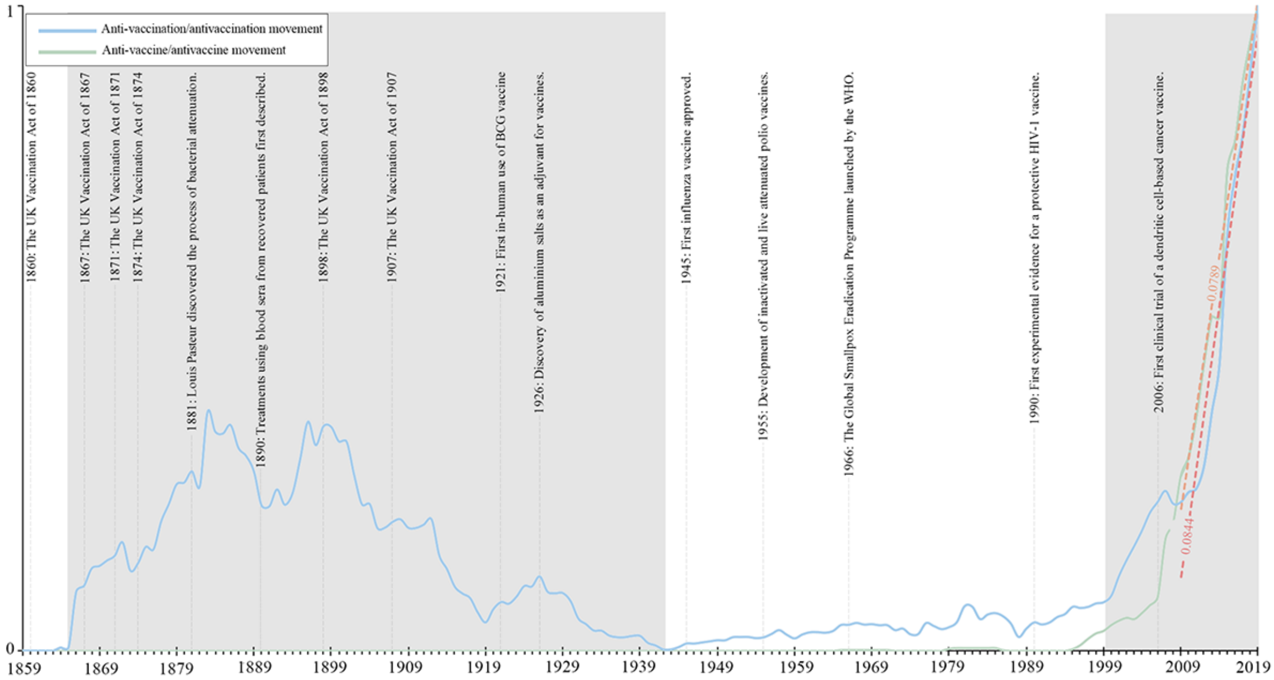


(a) (b)

Figure 5B shows four significant patterns in June, August, and September. The trigger events were #Chinese inactivated vaccine's Phase I/II trial was complete, and the participants had no adverse reactions# (June 16), #The chairman of Sinopharm said he was vaccinated and no adverse reactions# (August 18), #Oxford vaccine volunteers experienced adverse reactions# (September 9), and #The COVID-19 vaccine produced by Pfizer in America has side effects# (September 16), respectively. The topics were more active (ie, more messages posted) in September. From the perspective of emotion polarity, the messages' polarity about the side effects of COVID-19 vaccines can directly and positively affect people's emotion polarity. Over time, positive messages and negative messages became dominant in an alternating manner. In the end, Chinese views on side effects tended to be positive. The possible reason for this finding is that the positive emotions of prevailed messages on side effects result in herd behavior of followers [49]. After such information cascades, the population gradually accepted the existence of side effects and then reached a consensus (Figure 5B). Figure 3B shows similar patterns.

On the contrary, as *Nature News* reported, in high-income countries such as Europe, citizens' concerns no longer focus on the price but on safety [49]. Many people believe that vaccines will increase the immune system's burden, assuming that it exposes themselves to danger. In addition, a variety of other exaggerated rumors have spread on major social media outlets. According to Google Books Ngram Corpus (Figure 6), the co-occurrence of antivaccine and antivaccination movements has shown a clear upward trend in the past decades (the fitted slope of the antivaccination movement is 0.0844 and the antivaccine movement is 0.0789) [50]. The historical events that may have impacted the antivaccine and antivaccination movements in different eras are marked in Figure 6. It also shows that the concern for the safety of early vaccines has resulted in widespread protests. With the development of medical technology, vaccines have gradually become safer, and protests have decreased. Part of the reason for burgeoning movements in recent decades is due to ideological reasons [51]. However, as an exception, China has not experienced such scenarios. In view of this, in China, it is particularly important to collect public opinions via social media and improve policies in a timely manner.

Figure 6. Google Books Ngram Corpus facsimiles of word frequencies for "anti-vaccination/antivaccination movement" and "anti-vaccine/antivaccine movement" in the English corpus from 1859 to 2019. BCG: Bacillus Calmette–Guérin; WHO: World Health Organization.



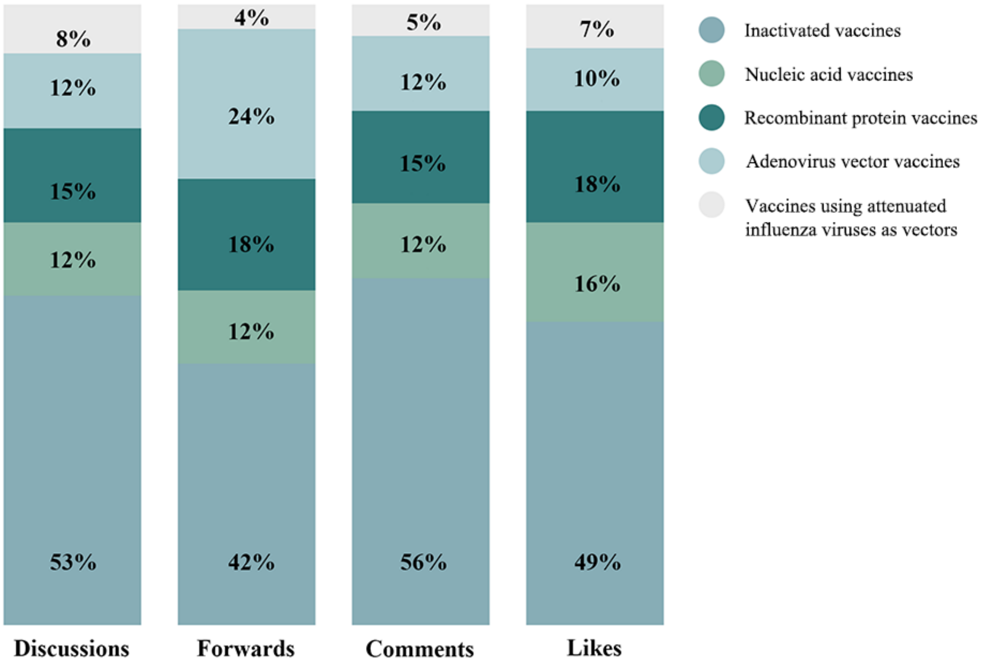
Hesitancy: Type of Vaccines

Vaccine hesitancy about COVID-19 vaccines is linked to populations who are reluctant or refuse to be vaccinated despite the availability of vaccination services [52]. The beliefs surrounding vaccine hesitancy are dynamic, complex, and context-specific, varying across time, place, and vaccine types, as well as complacency, convenience, and confidence. China is forging ahead in the race to develop COVID-19 vaccine candidates using five potential development routines: inactivated vaccines, adenovirus vector vaccines, vaccines using attenuated influenza viruses as vectors, recombinant protein vaccines, and nucleic acid vaccines (including RNA [ribonucleic acid] vaccines and DNA vaccines) [53]. China's National Medical Products Administration projected the country's production capacity of COVID-19 vaccines would reach 610 million doses annually by the end of 2020. Of these, the licensed vaccines for limited rollout would be inactivated vaccines [54].

Figure 7 shows the proportion of these five vaccines as mentioned by Chinese netizens. We have selected four dimensions—discussions, forwards, comments, and likes—to show the degree of attention on the different vaccines. Evidently, inactivated vaccines had substantially more discussions, forwards, comments, and likes than the other four vaccine types.

Compared with the other vaccines, inactivated vaccines are more acceptable to the public in China. The reason may be that the word "inactivated" in Chinese means "being killed" or "dead"; this gives people a sense of more security. For the same reason, the vaccines using attenuated influenza viruses as vectors, which means "alive influenza virus," has been rarely mentioned by netizens. It accounts for 8% (87/1117) of mentions (Figure 7 and Multimedia Appendix 3).

Figure 7. The distribution of netizens' attention to five different types of China-made vaccines.



Discussion

Principal Findings

Using rich random-sampling metadata of Weibo posts (more than 1.75 million messages with approximately 21.17 billion links), we conducted a country-specific study of real-time public awareness and behavioral responses to COVID-19 vaccines and vaccination from January to October 2020 in China. By studying the collective transmission behavior of Chinese netizens (domestic Chinese and those living abroad), this research revealed the paradigm shifts in public demand and events instigating public opinion in the context of the promotion of COVID-19 vaccination.

Beyond Affordability

Our findings strongly suggest that the Chinese public is divided in terms of vaccination prices and has differing expectations. Supporters accepted the current official price, whereas opponents claimed that it is unaffordable for their extended family. Arguably, the majority of people think that the price is lower than previously thought. If the Chinese government promises to pay for vaccination, taking 600 RMB per person as an example, the country would bear a financial burden of at least 840 billion RMB, considering its large population size. Therefore, introducing a policy for free COVID-19 vaccination is an extremely difficult decision for the Chinese government to make. If the government simply and directly included COVID-19 vaccines in medical insurance plans, it would impose a financial burden on hospitals, and the costs would eventually pass on to patients [55]. If commercial insurance is subsidized, another problem will arise: Chinese families will tend to insure children, who have a low risk of COVID-19, rather than high-risk elderly people [56,57].

Concerns Raised Over Efficacy

Our study is characterized by the striking feature that people's concerns about side effects tend to be positive over time. News of fast-tracked, China-made vaccine candidates are encouraging and helpful, but scientists urge caution [58]. Admittedly, the volunteers who received either the trial vaccine or placebo were low-risk, healthy adults rather than high-risk populations (eg, obese patients [59,60], patients with autoinflammatory syndrome [61], etc). Many scientists hold that there is uncertainty about the vaccine candidates' true efficacy when they are extended to the general populations [62]. More observations are needed to test the effectiveness of the vaccine [63,64]. On the contrary, the results of prevailing messages pinpoint that most Chinese people would accept the marginal risks (ie, side effects) with much confidence in the forthcoming vaccines.

Cognitive Dissonance Debunked

The literal meaning of the Chinese name for inactivated vaccines may provide a false sense of security. According to our survey, most Chinese people reached a consensus on the safety of inactivated vaccines due to cognitive dissonance. By contrast, people's expectations for inactivated vaccines were higher than other types of vaccines. There is no scientific evidence supporting the safety of inactivated vaccines compared to the other types. Furthermore, even when the vaccine is proven to be safe and effective, its acceptance will vary by group [65]. The World Health Organization has listed vaccine hesitancy as one of the top 10 threats to global health [66,67]. It is usually caused by the association of moral values in families [24,68], although this is unnecessary [69]. Nevertheless, the biggest problem in China is cognitive dissonance. Being echoed by daily communications, this finding indicates that this false cognitive predisposition may reinforce the tendency to vaccinate using inactivated vaccines and discourage vaccination via other types. Even worse, as vaccines lose their competitive advantages, people may not choose the most suitable one or even lose their right to choose. As rare events often attack the safety of vaccines [70], dissonance can lead to collective misbehavior.

Limitations

In this study, we retrieved more than 1.75 million Weibo messages in 108 languages; however, only messages in Simplified Chinese and Traditional Chinese were further investigated. These messages may come from domestic Chinese, those living abroad, or even foreign netizens living in China. Therefore, the findings reported in this study reflect COVID-19 vaccine acceptance in the Chinese-speaking population. However, each netizen account is authenticated according to the real-name verification policy of China, and the authenticity of the message can be guaranteed as they almost always come from real people rather than bots. With this in mind, this limitation does not undermine the significance of our findings.

The landscape of public opinion transmission is still ever-changing, especially in terms of price, side effects, etc. As the three major topics are still in the developing stage, a large number of messages and interactions showed up after study completion. However, the paradigm shifts of all the pertinent topics discussed here have been observed. Therefore, this limitation should not undermine the significance of the results either.

In addition, according to the 46th Statistical Report on Internet Development in China released by the China Internet Network Information Center in September 2020 [71] (Multimedia Appendix 4), only 10.3% of Chinese netizens are over 60 years of age. Therefore, to overcome the undersampling of the elderly population in this survey, the determinants of COVID-19 vaccine acceptance among older adults need to be further investigated [43].

Conclusions

At this critical moment in China, articulating the dynamic social paradigms of public engagement for COVID-19 vaccination is paramount for examining the practical strategies of social mobilization, wherein one sheds light on the other's significance. We scrutinized collective responses on COVID-19 vaccines and pertinent discourses in sociocultural paradigms to uncover collective propensities and consequences.

As an integral component of preparedness, the contextualized results reported in this study promise to provide illuminating benchmarks to bridge the gaps of health and risk communications. In China, the landscape of public opinion transmission on Chinese social media is unique, featuring a real-name verification policy. Therefore, the online collective propensities on COVID-19 vaccines and vaccination could resonantly echo daily responses in the real world, including those from domestic Chinese and those living abroad. Although there is no need for explicit antivaccine or antivaccination movements, the implicit channel of online public appeals is more vital than ever for improving policies.

The paradigms we identified to be the determinants of COVID-19 vaccine acceptance (eg, public appeals on affordability, efficacy, and preferences) could reframe a heuristic framework for extensive discussions, especially on vaccine-promoting policies in China. Reflecting on the unfolding findings, evidence communication is a heuristic way to nurture trustworthiness [72]. For instance, the government could consider using health insurance to balance state finances with individual expenditure. Making vaccine clinical trial data open and transparent is an effective way to assuage public skepticism. To eliminate strongly held but flawed cognitive predispositions, the government needs to increase the popularization of public science to communicate the side effects of drugs and strengthen publicity for all kinds of vaccines. Moreover, previous studies have indicated that once the vaccines are available, distributing them may become problematic [73-75]. The production capacity of vaccines has also been questioned by scholars. The impending worry is that new rounds of antivaccine and antivaccination movements will sprung up in a backlash of populism [58] and further undermine global efforts to curb the COVID-19 pandemic. Some scholars have turned their attention to disadvantaged groups and believe that the ethical framework needs to be improved to protect these groups' rights and interests during COVID-19 vaccination [75]. Arguably, affordable and effective vaccines offer hope for ending the pandemic, and open-minded and iterative policies fuel public engagement against the pandemic.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (grant 61801440); the Zhejiang Provincial Natural Science Foundation of China (grant LZ21F020004); the Fundamental Research Funds for the Central Universities; and the High-Quality and Cutting-Edge Disciplines Construction Project for Universities in Beijing (Internet Information, Communication University of China), State Key Laboratory of Media Convergence and Communication, and Communication University of China.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Typical topics related to COVID-19 vaccines, containing information about topic names (with Chinese translation), reading quantity, and date.

[DOCX File , 30 KB - jmir_v23i1e26089_app1.docx]

Multimedia Appendix 2
Price polarity table.
[DOCX File , 13 KB - jmir_v23i1e26089_app2.docx]

Multimedia Appendix 3
Discussions, forwards, comments, and likes related to China's five vaccine types.
[DOCX File , 16 KB - jmir_v23i1e26089_app3.docx]

Multimedia Appendix 4
Statistical Report on Internet Development in China.
[PDF File (Adobe PDF File), 7014 KB - jmir_v23i1e26089_app4.pdf]

References

1. Wadman M, You J. The vaccine wars. Science 2017 Apr 28;356(6336):364-365. [doi: 10.1126/science.356.6336.364] [Medline: 28450592]
2. Larson. Blocking information on COVID-19 can fuel the spread of misinformation. Nature 2020 Apr;580(7803):306. [doi: 10.1038/d41586-020-00920-w] [Medline: 32231320]
3. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]
4. Ten threats to global health in 2019. World Health Organization. 2019. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2020-10-28]
5. Kennedy M. Samoa arrests anti-vaccination activist as measles death toll rises. NPR. 2019 Dec 6. URL: https://www.npr.org/2019/12/06/785487606/samoa-arrests-anti-vaccination-activist-as-measles-death-toll-rises [accessed 2020-10-28]
6. Givetash L. Global measles cases surge amid stagnating vaccinations. NBC News. 2019 Dec 6. URL: http://www.nbcnews.com/news/amp/ncna1096921 [accessed 2020-10-28]
7. Bandari R, Zhou Z, Qian H, Tangherlini TR, Roychowdhury VP. A Resistant Strain: Revealing the Online Grassroots Rise of the Antivaccination Movement. Computer 2017 Nov;50(11):60-67. [doi: 10.1109/mc.2017.4041354]
8. Zadrozny B, Edwards E. Anti-vaccine groups take dangerous online harassment into the real world. NBC News. 2019 Dec 6. URL: https://www.nbcnews.com/news/amp/ncna1096461 [accessed 2020-10-28]
9. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
10. Depoux A, Martin S, Karafillakis E, Bsd RP, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 Apr;27(13):1-2. [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
11. Khan YH, Mallhi TH, Alotaibi NH, Alzarea AI, Alanazi AS, Tanveer N, et al. Threat of COVID-19 Vaccine Hesitancy in Pakistan: The Need for Measures to Neutralize Misleading Narratives. Am J Trop Med Hyg 2020 Aug;103(2):603-604 [FREE Full text] [doi: 10.4269/ajtmh.20-0654] [Medline: 32588810]
12. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro- and anti-vaccination views. Nature 2020 Jun;582(7811):230-233. [doi: 10.1038/s41586-020-2281-1] [Medline: 32499650]
13. Kennedy A, Lavail K, Nowak G, Basket M, Landry S. Confidence about vaccines in the United States: understanding parents' perceptions. Health Aff (Millwood) 2011 Jun;30(6):1151-1159 [FREE Full text] [doi: 10.1377/hlthaff.2011.0396] [Medline: 21653969]
14. Burton-Jeangros C, Golay M, Sudre P. Compliance and resistance to child vaccination: a study among Swiss mothers. Rev Epidemiol Sante Publique 2005 Sep;53(4):341-341. [doi: 10.1016/s0398-7620(05)84616-4] [Medline: 16353509]
15. Wu AC, Wisler-Sher DJ, Griswold K, Colson E, Shapiro ED, Holmboe ES, et al. Postpartum mothers' attitudes, knowledge, and trust regarding vaccination. Matern Child Health J 2008 Nov;12(6):766-773 [FREE Full text] [doi: 10.1007/s10995-007-0302-4] [Medline: 17987370]
16. Black S, Rappuoli R. A crisis of public confidence in vaccines. Sci Transl Med 2010 Dec 08;2(61):61mr1. [doi: 10.1126/scitranslmed.3001738] [Medline: 21148125]
17. Poland GA, Jacobson RM, Ovsyannikova IG. Trends affecting the future of vaccine development and delivery: the role of demographics, regulatory science, the anti-vaccine movement, and vaccinomics. Vaccine 2009 May 26;27(25-26):3240-3244 [FREE Full text] [doi: 10.1016/j.vaccine.2009.01.069] [Medline: 19200833]
18. MacDonald NE, Smith J, Appleton M. Risk perception, risk management and safety assessment: what can governments do to increase public confidence in their vaccine system? Biologicals 2012 Sep;40(5):384-388. [doi: 10.1016/j.biologicals.2011.08.001] [Medline: 21993306]
19. Leask J. Target the fence-sitters. Nature 2011 May 26;473(7348):443-445. [doi: 10.1038/473443a] [Medline: 21614055]

20. Leask J, Kinnersley P, Jackson C, Cheater F, Bedford H, Rowles G. Communicating with parents about vaccination: a framework for health professionals. BMC Pediatr 2012 Sep 21;12(1):154 [FREE Full text] [doi: 10.1186/1471-2431-12-154] [Medline: 22998654]
21. Callender D. Vaccine hesitancy: More than a movement. Hum Vaccin Immunother 2016 Sep;12(9):2464-2468 [FREE Full text] [doi: 10.1080/21645515.2016.1178434] [Medline: 27159558]
22. Carrieri V, Madio L, Principe F. Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. Health Econ 2019 Nov;28(11):1377-1382 [FREE Full text] [doi: 10.1002/hec.3937] [Medline: 31429153]
23. Wilson K, Atkinson K, Deeks S. Opportunities for utilizing new technologies to increase vaccine confidence. Expert Rev Vaccines 2014 Aug;13(8):969-977. [doi: 10.1586/14760584.2014.928208] [Medline: 24931799]
24. Amin AB, Bednarczyk RA, Ray CE, Melchiori KJ, Graham J, Huntsinger JR, et al. Association of moral values with vaccine hesitancy. Nat Hum Behav 2017 Dec;1(12):873-880. [doi: 10.1038/s41562-017-0256-5] [Medline: 31024188]
25. Larson HJ. Politics and public trust shape vaccine risk perceptions. Nat Hum Behav 2018 May;2(5):316. [doi: 10.1038/s41562-018-0331-6] [Medline: 30962595]
26. Martin B. Texas anti-vaxxers fear mandatory COVID-19 vaccines more than the virus itself. Texas Monthly. 2020 Mar 18. URL: https://www.texasmonthly.com/news/texas-anti-vaxxers-fear-mandatory-coronavirus-vaccines/ [accessed 2020-10-28]
27. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J, Lachapelle B. User-Driven Comments on a Facebook Advertisement Recruiting Canadian Parents in a Study on Immunization: Content Analysis. JMIR Public Health Surveill 2018 Sep 20;4(3):e10090 [FREE Full text] [doi: 10.2196/10090] [Medline: 30249585]
28. Larson HJ, Jarrett C, Eckersberger E, Smith DMD, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012. Vaccine 2014 Apr 17;32(19):2150-2159. [doi: 10.1016/j.vaccine.2014.01.081] [Medline: 24598724]
29. Geoghegan S, O'Callaghan KP, Offit PA. Vaccine Safety: Myths and Misinformation. Front Microbiol 2020 Mar 17;11:372 [FREE Full text] [doi: 10.3389/fmicb.2020.00372] [Medline: 32256465]
30. Plotkin SA, Offit PA, DeStefano F, Larson HJ, Arora NK, Zuber PLF, et al. The science of vaccine safety: Summary of meeting at Wellcome Trust. Vaccine 2020 Feb 18;38(8):1869-1880. [doi: 10.1016/j.vaccine.2020.01.024] [Medline: 31987690]
31. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 2020 Jul;26(7):1470-1477 [FREE Full text] [doi: 10.3201/eid2607.200282] [Medline: 32255761]
32. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res 2020 Jul;24:91-98 [FREE Full text] [doi: 10.1016/j.jare.2020.03.005] [Medline: 32257431]
33. Liu X, Zhang S. COVID-19: Face masks and human-to-human transmission. Influenza Other Respir Viruses 2020 Jul;14(4):472-473 [FREE Full text] [doi: 10.1111/irv.12740] [Medline: 32223072]
34. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020 Mar 05;382(10):970-971. [doi: 10.1056/NEJMc2001468] [Medline: 32003551]
35. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. N Engl J Med 2020 May 28;382(22):2158-2160 [FREE Full text] [doi: 10.1056/NEJMe2009758] [Medline: 32329972]
36. Clarke C, Prendecki M, Dhutia A, Ali MA, Sajjad H, Shivakumar O, et al. High Prevalence of Asymptomatic COVID-19 Infection in Hemodialysis Patients Detected Using Serologic Screening. J Am Soc Nephrol 2020 Sep;31(9):1969-1975. [doi: 10.1681/ASN.2020060827] [Medline: 32732391]
37. Haque A, Pant AB. Efforts at COVID-19 Vaccine Development: Challenges and Successes. Vaccines (Basel) 2020 Dec 06;8(4):739 [FREE Full text] [doi: 10.3390/vaccines8040739] [Medline: 33291245]
38. Wu S. Progress and Concept for COVID-19 Vaccine Development. Biotechnol J 2020 Jun;15(6):e2000147 [FREE Full text] [doi: 10.1002/biot.202000147] [Medline: 32304139]
39. Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. Nat Rev Drug Discov 2020 May;19(5):305-306. [doi: 10.1038/d41573-020-00073-5] [Medline: 32273591]
40. Ball P. Anti-vaccine movement could undermine efforts to end coronavirus pandemic, researchers warn. Nature 2020 May 13;581(7808):251. [doi: 10.1038/d41586-020-01423-4] [Medline: 32405043]
41. Leask J. Vaccines — lessons from three centuries of protest. Nature 2020 Sep 21;585(7826):499-501. [doi: 10.1038/d41586-020-02671-0]
42. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-10-28]
43. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2020 Oct 20:1-4 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
44. Weibo Report. Weibo. URL: https://data.weibo.com/report/reportDetail?id=43 [accessed 2020-10-28]
45. Rao Q, Zhang Z, Lv Y, Zhao Y, Bai L, Hou X. Factors Associated With Influential Health-Promoting Messages on Social Media: Content Analysis of Sina Weibo. JMIR Med Inform 2020 Oct 09;8(10):e20558 [FREE Full text] [doi: 10.2196/20558] [Medline: 33034569]

46. Yin F, Wu J, Shao X, Wu J. Topic reading dynamics of the Chinese Sina-Microblog. Chaos, Solitons & Fractals: X 2020 Mar;5:100031. [doi: 10.1016/j.csfx.2020.100031]
47. Tuljapurkar S. The mathematics of infection. Science 1991 Oct 25;254(5031):591-592. [doi: 10.1126/science.254.5031.591] [Medline: 17806979]
48. Liu Q, Li S. Word similarity computing based on HowNet. International Journal of Computational Linguistics & Chinese Language Processing 2002;7(2):59-76 [FREE Full text]
49. Tan X, Cousins K. Herding behavior in social media networks in China. In: AMCIS. 2016 Presented at: 22nd Americas Conference on Information Systems; August 11-14, 2016; San Diego, CA URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1065&context=amcis2016
50. Michel J, Shen YK, Aiden AP, Veres A, Gray MK, Google Books Team, et al. Quantitative analysis of culture using millions of digitized books. Science 2011 Jan 14;331(6014):176-182 [FREE Full text] [doi: 10.1126/science.1199644] [Medline: 21163965]
51. Rutjens BT, Sutton RM, van der Lee R. Not All Skepticism Is Equal: Exploring the Ideological Antecedents of Science Acceptance and Rejection. Pers Soc Psychol Bull 2018 Mar;44(3):384-405 [FREE Full text] [doi: 10.1177/0146167217741314] [Medline: 29191107]
52. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert Rev Vaccines 2015 Jan;14(1):99-117. [doi: 10.1586/14760584.2015.964212] [Medline: 25373435]
53. China approves inactivated COVID-19 vaccines for clinical trials. People's Daily Online. 2020 Apr 14. URL: http://en.people.cn/n3/2020/0414/c90000-9679309.html [accessed 2020-10-28]
54. China's COVID-19 vaccine capacity to reach 610 million. Global Times. 2020 Oct 21. URL: https://www.globaltimes.cn/content/1204143.shtml [accessed 2020-10-28]
55. Luan M, Shao X, Dou F. Financial conditions, health care provision, and patient outcomes: Evidence from Chinese public hospitals. Economics Letters 2020 Jan;186:108876. [doi: 10.1016/j.econlet.2019.108876]
56. Li G, Li Z, Lv X. The ageing population, dependency burdens and household commercial insurance purchase: evidence from China. Applied Economics Letters 2020 Apr 14:1-5. [doi: 10.1080/13504851.2020.1752360]
57. Bruine de Bruin W. Age differences in COVID-19 risk perceptions and mental health: Evidence from a national US survey conducted in March 2020. The Journals of Gerontology: Series B 2020 May 29 [FREE Full text] [doi: 10.1093/geronb/gbaa074] [Medline: 32470120]
58. Torreele E. The rush to create a covid-19 vaccine may do more harm than good. BMJ 2020 Aug 18;370:m3209. [doi: 10.1136/bmj.m3209] [Medline: 32816760]
59. Kwok S, Adam S, Ho JH, Iqbal Z, Turkington P, Razvi S, et al. Obesity: A critical risk factor in the COVID-19 pandemic. Clin Obes 2020 Dec;10(6):e12403 [FREE Full text] [doi: 10.1111/cob.12403] [Medline: 32857454]
60. Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, et al. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. Obes Rev 2020 Nov;21(11):e13128 [FREE Full text] [doi: 10.1111/obr.13128] [Medline: 32845580]
61. Koirala A, Joo YJ, Khatami A, Chiu C, Britton PN. Vaccines for COVID-19: The current state of play. Paediatr Respir Rev 2020 Sep;35:43-49 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.010] [Medline: 32653463]
62. de Vrieze J. Reinfections, still rare, provide clues on immunity. Science 2020 Nov 20;370(6519):895-897. [doi: 10.1126/science.370.6519.895] [Medline: 33214256]
63. Callaway E. Why Oxford's positive COVID vaccine results are puzzling scientists. Nature 2020 Dec 23;588(7836):16-18. [doi: 10.1038/d41586-020-03326-w] [Medline: 33230278]
64. Cyranoski D. Why emergency COVID-vaccine approvals pose a dilemma for scientists. Nature 2020 Dec 23;588(7836):18-19. [doi: 10.1038/d41586-020-03219-y] [Medline: 33230275]
65. Callaghan T, Moghtaderi A, Lueck JA, Hotez PJ, Strych U, Dor A, et al. Correlates and Disparities of COVID-19 Vaccine Hesitancy. SSRN Journal Preprint posted online Aug 12, 2020. [doi: 10.2139/ssrn.3667971]
66. Time to vaccinate against hesitancy. Nat Med 2019 Jul 3;25(7):1023-1023. [doi: 10.1038/s41591-019-0524-1] [Medline: 31270495]
67. The COVID vaccine challenges that lie ahead. Nature 2020 Nov 24;587(7835):522-522. [doi: 10.1038/d41586-020-03334-w] [Medline: 33235368]
68. Kupferschmidt K. Can skeptical parents be persuaded to vaccinate? Science 2017 Apr 28. [doi: 10.1126/science.aal1108]
69. Eyal N. Unnecessary hesitancy on human vaccine tests. Science 2020 Jul 10;369(6500):150-151. [doi: 10.1126/science.abc8264] [Medline: 32646991]
70. Kwok R. Vaccines: The real issues in vaccine safety. Nature 2011 May 26;473(7348):436-438. [doi: 10.1038/473436a] [Medline: 21614053]
71. Statistical Report on Internet Development in China. China Internet Network Information Center. 2020 Sep. URL: https://cnnic.com.cn/IDR/ReportDownloads/202012/P020201201530023411644.pdf [accessed 2020-12-17]
72. Blastland M, Freeman ALJ, van der Linden S, Marteau TM, Spiegelhalter D. Five rules for evidence communication. Nature 2020 Nov 18;587(7834):362-364. [doi: 10.1038/d41586-020-03189-1] [Medline: 33208954]

73. Roope LSJ, Buckell J, Becker F, Candio P, Violato M, Sindelar JL, et al. How Should a Safe and Effective COVID-19 Vaccine be Allocated? Health Economists Need to be Ready to Take the Baton. Pharmacoecon Open 2020 Dec;4(4):557-561 [FREE Full text] [doi: 10.1007/s41669-020-00228-5] [Medline: 32880878]
74. Mosley M. Be Hopeful but Cautious about COVID-19 Vaccine. Emergency Medicine News 2020;42(9):2-8 [FREE Full text] [doi: 10.1097/01.EEM.0000697768.81268.6b]
75. Liu Y, Salwi S, Drolet BC. Multivalue ethical framework for fair global allocation of a COVID-19 vaccine. J Med Ethics 2020 Aug;46(8):499-501 [FREE Full text] [doi: 10.1136/medethics-2020-106516] [Medline: 32532826]

Abbreviations

AE: adverse event
R_0: reproduction ratio
RNA: ribonucleic acid
SRS/I: susceptible, reading, susceptible/immunized

Edited by G Eysenbach; submitted 27.11.20; peer-reviewed by H Wang, A Chang, Z Su; comments to author 11.12.20; revised version received 22.12.20; accepted 24.12.20; published 15.01.21.

Please cite as:
Yin F, Wu Z, Xia X, Ji M, Wang Y, Hu Z
Unfolding the Determinants of COVID-19 Vaccine Acceptance in China
J Med Internet Res 2021;23(1):e26089
URL: http://www.jmir.org/2021/1/e26089/
doi:10.2196/26089
PMID:33400682

Original Paper

Maternal Mental Health Status and Approaches for Accessing Antenatal Care Information During the COVID-19 Epidemic in China: Cross-Sectional Study

Hong Jiang[1,2,3,4*], MD, PhD; Longmei Jin[4,5*], MD; Xu Qian[1,2,3,4], MD, PhD; Xu Xiong[6], DrPH; Xuena La[1,2], MD; Weiyi Chen[1,2], MD; Xiaoguang Yang[1,2,3,4], PhD; Fengyun Yang[7], MD; Xinwen Zhang[8], MD; Nazhakaiti Abudukelimu[1]; Xingying Li[1]; Zhenyu Xie[9], MD; Xiaoling Zhu[10], MD; Xiaohua Zhang[4,5], MD; Lifeng Zhang[7], MD; Li Wang[11], MD, PhD; Lingling Li[12], MPH; Mu Li[13,14], MD, PhD

[1]School of Public Health, Fudan University, Shanghai, China
[2]Key Lab of Health Technology Assessment (National Health Commission), Fudan University, Shanghai, China
[3]Global Health Institute, Fudan University, Shanghai, China
[4]Minhang Branch, School of Public Health, Fudan University, Shanghai, China
[5]Minhang Maternal and Child Health Care Hospital, Shanghai, China
[6]School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, United States
[7]Jiading Maternal and Child Health Care Hospital, Shanghai, China
[8]The Fourth People’s Hospital of Shaanxi Province, Xi’an, China
[9]Pudong New District Maternal and Child Health Care Hospital, Shanghai, China
[10]Leping Maternal and Child Health Care Hospital, Leping, China
[11]Changzhou Maternal and Child Health Care Hospital, Changzhou, China
[12]Changzheng Hospital, Second Military Medical University, Shanghai, China
[13]School of Public Health, The University of Sydney, Sydney, Australia
[14]China Studies Centre, The University of Sydney, Sydney, Australia
*these authors contributed equally

Corresponding Author:
Hong Jiang, MD, PhD
School of Public Health
Fudan University
Mailbox 175
138 Yixueyuan Road
Shanghai, 200032
China
Phone: 86 2154237267
Email: h_jiang@fudan.edu.cn

Abstract

Background: China was the first country in the world to experience a large-scale COVID-19 outbreak. The rapid spread of the disease and enforcement of public health measures has caused distress among vulnerable populations such as pregnant women. With a limited understanding of the novel, emerging infectious disease, pregnant women have sought ways to access timely and trusted health care information. The mental health status of pregnant women during this public health emergency, as well as how they responded to the situation and where and how they obtained antenatal care information, remain to be understood.

Objective: This study aimed to evaluate the mental health status of pregnant women during the COVID-19 epidemic in China by measuring their perceived stress, anxiety, and depression levels; explore the approaches used by them to access antenatal health care information; and determine their associations with maternal mental health status.

Methods: We conducted a web-based, cross-sectional survey to assess the mental health status of Chinese pregnant women by using the validated, Chinese version of Perceived Stress Scale, Self-Rating Anxiety Scale, and Edinburgh Depression Scale. We also collected information on the various approaches these women used to access antenatal care information during the early stage of the COVID-19 epidemic, from February 5 to 28, 2020.

Results: A total of 1873 pregnant women from 22 provinces or regions of China participated in the survey. The prevalence of perceived stress, anxiety, and depression among these participants was 89.1% (1668/1873; 95% CI 87.6%, 90.4%), 18.1% (339/1873; 95% CI 16.4%, 19.9%), and 45.9% (859/1873; 95% CI 43.6%, 48.1%), respectively. Hospitals' official accounts on the Chinese social media platforms WeChat and Weibo were the most popular channels among these pregnant women to obtain antenatal care information during the COVID-19 outbreak. Access to antenatal care information via the hospitals' official social media accounts was found to be associated with a significantly lower risk of perceived stress (adjusted odds ratio [aOR] 0.46, 95% CI 0.30-0.72; *P*=.001), anxiety (aOR 0.53, 95% CI 0.41-0.68; *P*<.001), and depression (aOR 0.73, 95% CI 0.59-0.91; *P*=.005). Access to health care information via hospital hotlines or SMS was found to be significantly associated with a lower risk of anxiety only (OR 0.77, 95% CI 0.60-0.98; *P*=.04).

Conclusions: During the COVID-19 outbreak in China, pregnant women experienced high levels of perceived stress, anxiety, and depression. During such public health emergencies, mental health care services should be strengthened to reassure and support pregnant women. Specific information targeted at pregnant women, including information on how to cope in an emergency or major disease outbreak, developed and disseminated by health care institutions via social media platforms could be an effective way to mitigate mental health challenges and ensure epidemic preparedness and response in the future.

(J Med Internet Res 2021;23(1):e18722) doi:10.2196/18722

KEYWORDS

COVID-19; mental health; perceived stress; anxiety; depression; antenatal care information; social media platform; pregnancy; women

Introduction

COVID-19 first emerged in Wuhan city in China in December 2019 [1,2]. Soon after, the Chinese government initiated the level-one (ie, the highest level) response to major public health emergencies [3], and the World Health Organization declared the COVID-19 outbreak as a Public Health Emergency of International Concern on January 30, 2020. By February 29, 2020, the number of confirmed COVID-19 cases in China reached over 79,000, with 2800 deaths reported and 20% of the cases classified as severe [4]. March onward, the COVID-19 outbreak spread to almost all nations worldwide and evolved into a full-blown pandemic [5]. By the end of August 2020, more than 28 million confirmed COVID-19 cases and over 900,000 deaths had been reported worldwide [6]. The Chinese government implemented timely measures, including city lockdown, self-quarantine, and social distancing across the nation to contain the transmission of COVID-19. People were advised to stay at home and avoid going to public places and social gatherings [7]. The rapid spread of the epidemic and the consequent enforcement of public health measures had unavoidably raised fear and panic reaction among the public [8-10].

Pregnant women and their unborn children are especially vulnerable to major epidemics or natural disasters [11,12]. During the early stage of the COVID-19 epidemic in China, with limited information and understanding about the new emerging infectious disease, pregnant women looked for different ways to access timely and trusted information regarding the causes, possible routes of transmission, prevention, and self-protection approaches to protect themselves and their unborn children [13]. Given the strict public health emergency measures such as travel restriction and social distancing, pregnant women had been isolated from their professional and social support, including hospitals that provided routine antenatal care services. Such interrupted contact with health institutions could exacerbate uncertainty among these women about service provision and their own health status. Routine maternal health services have often been affected during major infectious disease outbreaks in the past. For example, during the 2003 SARS outbreak, all antenatal services in Hong Kong, China, which were considered nonessential were suspended, and women were discharged as early as possible after childbirth [14]. Since the announcement of the level-one response to major public health emergencies was made in China in late January 2020 for COVID-19, some hospitals focused on containing the spread of COVID-19, and obstetric-related services were suspended as a result [15]. Furthermore, procedures of accessing maternal service were modified; for instance, most hospitals required prior online registration and some hospitals limited the number of family members accompanying the pregnant woman to her antenatal clinic appointment to a maximum of one person[16]. A previous review has shown that during infectious disease outbreaks, pregnant women are likely to experience psychological distress; they are often concerned about the wellbeing of their fetus, when faced with the decision of whether to comply with the recommended prevention and treatment guidelines with their effectiveness and potential side effects yet to be understood, and the disrupted routines of maternal health services [16]. Thus, timely access to information about antenatal care services and education about public health challenges, as well as knowledge of self-protection from trusted sources, could play important roles in reassuring and supporting pregnant women during a public health emergency.

Like the popularity of Facebook and WhatsApp in many countries, a number of homegrown social media platforms such as WeChat and Weibo have penetrated the daily lives of Chinese people. WeChat is now the most popular mobile phone app in China with over 654 million users as of 2019, followed by Weibo with nearly 360 million users as of 2019 [17]. An increasing number of private enterprises, government organizations, and hospitals have established WeChat or Weibo official accounts (WOAs) to disseminate service information and promote interaction with end users. WeChat and Weibo are also the two most popular social media platforms used by hospitals to inform the public about the services they offered and provide health education and counselling free of charge.

China was the first county in the world to experience the COVID-19 outbreak. Understanding how the epidemic and public health measures, including restricted mobility, affected maternal mental health and whether the ways of accessing antenatal care information affected the mental wellbeing of pregnant women would provide important scientific information to develop effective support strategies and preparatory mechanisms for future global public health emergencies. Therefore, the aims of this study were to (1) determine the mental health status of pregnant women during the early stage of the COVID-19 epidemic in China, by measuring their perceived stress, anxiety, and depression levels; (2) explore the various approaches used by them to access antenatal care information; and (3) determine their associations with the maternal mental health status.

Methods

Study Design

We conducted a web-based, cross-sectional survey from February 5 to 28, 2020, during the peak of the COVID-19 epidemic in China. The questionnaire was designed using the online survey tool Wenjuanxing platform [18]. The contents of the survey gathered participants' demographic characteristics and evaluated their COVID-19 self-protection behaviors, knowledge of antenatal COVID-19–related care, mental health status, and channels of obtaining antenatal care information. The study was approved by the Institutional Review Board of School of Public Health, Fudan University, Shanghai, China (IRB#2020-02-0803).

Data Collection

We used direct online and snowball recruitment methods through our research collaboration network in an attempt to achieve higher participation from different regions of China. We sent the quick response (QR) code or the link to the survey questionnaire to the heads of maternal and child health (MCH) departments of MCH hospitals and comprehensive hospitals; these included 2 district MCH hospitals in Shanghai and 1 municipal MCH hospital in Changzhou, Jiangsu Province; 1 county MCH hospital in Leping, Jiangxi Province; and 1 municipal comprehensive hospital in Xi'an, Shaanxi Province. These department heads sent the electronic QR code or the survey URL to the staff of their respective departments and posted the recruitment details on the walls of the antenatal clinics. Pregnant women who visited the antenatal clinics in these facilities during the survey period were invited to participate in the survey; they were required to either scan the QR code or access the URL via a mobile phone or other digital devices. On the first page of the survey, a brief introduction to the survey, including the purpose and contents of the web-based questionnaire and the estimated time to complete the survey, was provided. At this point, participants could either choose to continue to answer the questionnaire or close the page to exit the survey. Participants were also encouraged to share the survey QR code or URL with other pregnant women in their network. Upon clicking the "submit" button at survey completion, key antenatal care knowledge relevant to COVID-19 (which were also the answers to the questions) would be displayed on the screen.

Survey Design

Participants' sociodemographic and obstetric information, including age, education level, status of employment, parity, gestational age, and pregnancy complication, were collected. The survey comprised 10 questions regarding personal self-protection behaviors to prevent COVID-19 and knowledge of antenatal care required during the COVID-19 epidemic, based on the joint Chinese professional societies' guidelines [19]. The 4 behavior questions, including 2 multiple choice questions, were about (1) avoiding stepping out of the home, (2) wearing a face mask in public places, (3) practicing hand sanitization, and (4) practicing other protection behaviors such as avoiding using the public transport and avoiding contact with wild animals and consumption of wild animal meat. The 6 antenatal care questions, including 3 multiple choice questions, evaluate participant's knowledge of (1) whether the antenatal health visit should be cancelled, (2) whether pregnant women could undergo a computed tomography scan, (3) symptoms urging pregnant women to visit the hospital, (4) where to seek care in case of fever, (5) should newborn babies be quarantined if the mother is diagnosed with COVID-19, and (6) whether breastfeeding is recommended if the mother has COVID-19 (see Multimedia Appendix 1). Each question was assigned an individual score. The total score of self-protection behaviors was 4 and that of antenatal care knowledge was 6. Survey participants were categorized into the following 2 groups if their total scores were above the median score: better self-protection behavior group and higher antenatal care knowledge group.

Mental Well-Being Assessment

Perceived Stress

We used the validated Chinese Perceived Stress Scale (CPSS) to assess perceived stress among pregnant women [20]. The CPSS is a 14-item, self-reported questionnaire with 7 positive and 7 negative items measuring the degree to which individuals appraise situations in their lives as stressful [21,22]. Each item is scored on a 5-point Likert scale ranging from 1 ("never") to 5 ("very often") [22]. The total score of this scale is computed by summing the scores of all positive items and reverse-scoring all negative items. The cutoff point for CPSS for the Chinese population is 25. Therefore, participants who scored higher than 25 were considered to be experiencing perceived stress [22-24].

Anxiety

The validated, 20-item Chinese version of the Self-Rating Anxiety Scale (SAS) was used to measure participants' anxiety symptoms [25]. Each item on the SAS is scored using a 4-point Likert scale, with scores ranging from 1 ("a little of the time") to 4 ("most of the time"). Items 5, 9, 13, 17, and 19 are negatively keyed. The total score is computed by multiplying the sum of the scores of positive items and the reverse-scores of negative items by 1.25. The total score of SAS ranges from 20 to 80 [26]. Participants scoring 50 and above were considered to be experiencing anxiety [27,28].

Depression

The 10-item Chinese version of the Edinburgh Depression Scale (EDS) was used to evaluate maternal depression [29]. Each item was scored on a 4-point Likert scale, with scores ranging from 0 to 3. The total score ranges from 0 to 30, and the cut-off scores range from 9 to 13. For this study, we considered a cut-off score of 9. Participants with a score of 10 and above were considered to be experiencing depression [30-32].

Accessing Antenatal Care Information

The question on how participants accessed antenatal care information during the COVID-19 epidemic was in a multiple-choice format: hospitals' hotline, mobile phone SMS, hospitals' official WeChat accounts, hospitals' official Weibo accounts, WeChat or Weibo moments (an information sharing forum) posted by friends and family members, and digital message or verbal advice shared by friends and family members. Participants who selected either the hospitals' official WeChat or Weibo accounts were categorized as accessing information via hospitals' official accounts on social media platforms. Those who selected WeChat or Weibo moments from friends and family members or other means involving friends and family members were categorized as accessing information via friends and family members.

Accessing antenatal care information via hotlines or SMS represents accessing antenatal care information from a reliable source and using a traditional approach. Accessing health care information via hospital official accounts on social media represents accessing antenatal care information via a reliable source and using a social media approach. Accessing health care information via friends or family members represents accessing information from nonprofessional sources. Analysis of these 3 different approaches will help clarify the roles of both the information sources and access channels.

Quality Control

For the purpose of quality control of the web-based survey, we incorporated 1 question after the CPSS scale: "The purpose of this question is to verify if a participant has answered the question carefully. Please select 'always' as you are told." If a participant failed to answer the question as required, their questionnaire response would be regarded as invalid. Among the 2186 pregnant women who participated in the survey, 1873 returned a valid questionnaire.

Data Analysis

We evaluated the mean (SD) values for normally distributed continuous variables and median (P25-P75) values for non-normally distributed variables. Categorical variables were described in proportions. Chi-square test was used to analyze the categorical variables, and multiple binary logistic regression analyses were performed to examine the associations between different approaches of accessing antenatal care information and stress, anxiety, and depression after adjusting for social economic factors such as age; education; employment status; living in urban, suburban, or rural areas; province or region of current residence; obstetric conditions (eg, parity, trimester, and presence of complications); score of COVID-19 prevention self-protection behaviors; and score of COVID-19–related antenatal care knowledge. The significance level was set at $P<.05$, and data were analyzed using SPSS software for Windows (version 17.0; SPSS Inc.) and R Statistical Software (version 3.6.3).

Results

The mean age of the 1873 participants was 29 (SD 4.10) years. The distribution of women's current residence covered 22 provinces or regions of China, with the majority of participants based in Shanghai (1415/1873, 75.5%) and a small proportion based in Hubei Province (23/1873, 1.2%)—the most severely affected area due to the COVID-19 outbreak. The majority of women had completed college and above level of education, were employed, and primiparous. The number of pregnant women in their first, second, and third trimesters was almost evenly distributed. The median score of COVID-19 self-protection behavior was 3, with individual scores ranging from 1 to 4. The median score of antenatal care knowledge was 4, with individual scores ranging from 0 to 6 (Table 1).

Table 1. Characteristics of pregnant women and their approaches of accessing antenatal care information during the COVID-19 outbreak in China (N=1873).

Characteristic	Accessing antenatal care information via hospital hotlines or SMS				Accessing health care information via hospital official accounts on social media platforms				Accessing health care information via friends or family members			
	No, n (%)[a]	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]
Overall	988 (52.7)	885 (47.3)	—[c]	—	457 (24.4%)	1416 (75.6)	—	—	1127 (60.2)	746 (39.8)	—	—
Age (years)[d]			0.9 (1)	.34			4.8 (1)	.03			4.6 (1)	.03
<29	373 (51.4)	353 (48.6)			197 (27.1)	529 (72.9)			459 (63.2)	267 (36.8)		
≥29	615 (53.6)	532 (46.4)			260 (22.7)	887 (77.3)			668 (58.2)	479 (41.8)		
Education level			10.7 (2)	.005			51.8 (2)	<.001			7.8 (2)	.02
Junior high or lower	103 (46)	121 (54)			88 (39.3)	136 (60.7)			143 (63.8)	81 (36.2)		
Senior high	130 (46.9)	147 (53.1)			92 (33.2)	185 (66.8)			184 (66.4)	93 (33.6)		
College or higher	755 (55)	617 (45)			277 (20.2)	1095 (79.8)			800 (58.3)	572 (41.7)		
Employment status			1.1 (1)	.30			8.8 (1)	.003			2.3 (1)	.13
Unemployed	140 (55.8)	111 (44.2)			80 (31.9)	171 (68.1)			162 (64.5)	89 (35.5)		
Employed	848 (52.3)	774 (47.7)			377 (23.2)	1245 (76.8)			965 (59.5)	657 (40.5)		
Parity			17.9 (1)	<.001			1.5 (1)	.23			3.1 (1)	.08
Primiparous	685 (56.3)	531 (43.7)			286 (23.5)	930 (76.5)			714 (58.7)	502 (41.3)		
Multiparous	303 (46.1)	354 (53.9)			171 (26)	486 (74)			413 (62.9)	244 (37.1)		
Trimester			21.1 (2)	<.001			18.6 (2)	<.001			3.2 (2)	.20
1st	271 (45.3)	327 (54.7)			112 (18.7)	486 (81.3)			348 (58.2)	250 (41.8)		
2nd	384 (54.6)	319 (45.4)			204 (29)	499 (71)			441 (62.7)	262 (37.3)		
3rd	333 (58.2)	239 (41.8)			141 (24.7)	431 (75.3)			338 (59.1)	234 (40.9)		
Living area			0.4 (2)	.83			36.0 (2)	<.001			4.8 (2)	.09
Urban	699 (52.5)	632 (47.5)			286 (21.5)	1045 (78.5)			781 (58.7)	550 (41.3)		
Suburban	174 (52.4)	158 (47.6)			86 (25.9)	246 (74.1)			208 (62.7)	124 (37.3)		
Rural	115 (54.8)	95 (45.2)			85 (40.5)	125 (59.5)			138 (65.7)	72 (34.3)		
Current residence			21.8 (1)	<.001			7.1 (1)	.008			0.0 (1)	.96
Non-Shanghai	285 (62.2)	173 (37.8)			133 (29)	325 (71)			276 (60.3)	182 (39.7)		
Shanghai	703 (49.7)	712 (50.3)			324 (22.9)	1091 (77.1)			851 (60.1)	564 (39.9)		
Pregnancy complications			0.4 (1)	.53			0.8 (1)	.38			0.1 (1)	.75

Characteristic	Accessing antenatal care information via hospital hotlines or SMS				Accessing health care information via hospital official accounts on social media platforms				Accessing health care information via friends or family members			
	No, n (%)[a]	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]
No	737 (53.2)	649 (46.8)			331 (23.9)	1055 (76.1)			831 (60)	555 (40)		
Yes	251 (51.5)	236 (48.5)			126 (25.9)	361 (74.1)			296 (60.8)	191 (39.2)		
Median score of COVID-19 prevention self-protection behavior			10.8 (1)	.001			4.3 (1)	.04			9.8 (1)	.002
Low	784 (54.9)	645 (45.1)			365 (25.5)	1064 (74.5)			888 (62.1)	541 (37.9)		
High	204 (45.9)	240 (54.1)			92 (20.7)	352 (79.3)			239 (53.8)	205 (46.2)		
Median score of COVID-19 antenatal care knowledge			2.0 (1)	.16			7.3 (1)	.007			3.5 (1)	.06
Low	722 (51.8)	672 (48.2)			362 (26)	1032 (74)			856 (61.4)	538 (38.6)		
High	266 (55.5)	213 (44.5)			95 (19.8)	384 (80.2)			271 (56.6)	208 (43.4)		

[a] Percentages reported parenthetically are based on the sum of absolute values of each subcategory (or row) listed under “Characteristic.”

[b] Chi-square test.

[c] Data not available.

[d] The mean age of 29 years was used to categorize participants into the 2 groups: ≥29 and <29 years.

The mean score of perceived stress was 35.21 (SD 7.58), as measured by CPSS. The median scores of anxiety and depression as measured by SAS and EDS were 42.50 (IQR 37.50-47.50) and 9.00 (IQR 7.00, 11.00), respectively. The prevalence of perceived stress, anxiety, and depression symptoms among the participants was reported to be 89.1% (1668/1873; 95% CI 87.6%, 90.4%), 18.1% (339/1873; 95% CI 16.4%, 19.9%), and 45.9% (859/1873; 95% CI 43.6%, 48.1%), respectively. Univariate analysis showed that pregnant women who were employed, lived in Shanghai, had completed higher education, and had scored higher for COVID-19 prevention knowledge and self-protection behaviors showed a negative association for experiencing at least one form of the 3 mental health disorders assessed. On the other hand, pregnant women who were multiparous, resided in rural areas, and had pregnancy complications were more likely to experience at least one form of the 3 mental health disorders assessed, compared to those who were primiparous, lived in urban areas, and did not have pregnancy complications (Table 2).

Table 2. Characteristics of pregnant women and their perceived stress, anxiety, and depression levels (N=1873).

Characteristic	Perceived stress				Anxiety				Depression			
	No, n (%)[a]	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]
Overall	205 (10.9)	1668 (89.1)	—[c]	—	1534 (81.9)	339 (18.1)	—	—	1014 (54.1)	859 (45.9)	—	—
Age (years)[d]			2.5 (1)	.11			1.4 (1)	.24			0.1 (1)	.70
<29	69 (9.5)	657 (90.5)			585 (80.6)	141 (19.4)			389 (53.6)	337 (46.4)		
≥29	136 (11.9)	1011 (88.1)			949 (82.7)	198 (17.3)			625 (54.5)	522 (45.5)		
Education			17.1 (2)	<.001			28.9 (2)	<.001			22.5 (2)	<.001
Junior high or below	10 (4.5)	214 (95.5)			159 (71)	65 (29)			102 (45.5)	122 (54.5)		
Senior high	21 (7.6)	256 (92.4)			214 (77.3)	63 (22.7)			124 (44.8)	153 (55.2)		
College or above	174 (12.7)	1198 (87.3)			1161 (84.6)	211 (15.4)			788 (57.4)	584 (42.6)		
Employment status			4.2 (1)	.04			2.3 (1)	.13			3.1 (1)	.08
Unemployed	18 (7.2)	233 (92.8)			197 (78.5)	54 (21.5)			123 (49)	128 (51)		
Employed	187 (11.5)	1435 (88.5)			1337 (82.4)	285 (17.6)			891 (54.9)	731 (45.1)		
Parity			0.6 (1)	.45			2.3 (1)	.13			5.8 (1)	.02
Primiparous	138 (11.3)	1078 (88.7)			1008 (82.9)	208 (17.1)			683 (56.2)	533 (43.8)		
Multiparous	67 (10.2)	590 (89.8)			526 (80.1)	131 (19.9)			331 (50.4)	326 (49.6)		
Trimester			2.9 (2)	.23			1.2 (2)	.56			0.5 (2)	.78
1st	75 (12.5)	523 (87.5)			490 (81.9)	108 (18.1)			330 (55.2)	268 (44.8)		
2nd	76 (10.8)	627 (89.2)			583 (82.9)	120 (17.1)			380 (54.1)	323 (45.9)		
3rd	54 (9.4)	518 (90.6)			461 (80.6)	111 (19.4)			304 (53.1)	268 (46.9)		
Living area			9.3 (2)	.01			8.2 (2)	.02			9.8 (2)	.008
Urban	157 (11.8)	1174 (88.2)			1103 (82.9)	228 (17.1)			743 (55.8)	588 (44.2)		
Suburban	38 (11.4)	294 (88.6)			274 (82.5)	58 (17.5)			178 (53.6)	154 (46.4)		
Rural	10 (4.8)	200 (95.2)			157 (74.8)	53 (25.2)			93 (44.3)	117 (55.7)		
Current residence			17.3 (1)	<.001			7.9 (1)	.005			5.1 (1)	.02
Non-Shanghai	26 (5.7)	432 (94.3)			355 (77.5)	103 (22.5)			227 (49.6)	231 (50.4)		
Shanghai	179 (12.7)	1236 (87.3)			1179 (83.3)	236 (16.7)			787 (55.6)	628 (44.4)		
Pregnancy complications			2.5 (1)	.12			5.3 (1)	.02			1.8 (1)	.18
No	161 (11.6)	1225 (88.4)			1152 (83.1)	234 (16.9)			763 (55.1)	623 (44.9)		

Characteristic	Perceived stress				Anxiety				Depression			
	No, n (%)[a]	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]	No, n (%)	Yes, n (%)	Chi-square (*df*)	*P* value[b]
Yes	44 (9)	443 (91)			382 (78.4)	105 (21.6)			251 (51.5)	236 (48.5)		
Score of COVID-19 prevention self-protection behaviors			1.2 (1)	.27			0.6 (1)	.45			0.3 (1)	.61
Low	150 (10.5)	1279 (89.5)			1165 (81.5)	264 (18.5)			769 (53.8)	660 (46.2)		
High	55 (12.4)	389 (87.6)			369 (83.1)	75 (16.9)			245 (55.2)	199 (44.8)		
Score of COVID-19 antenatal care knowledge			11.0 (1)	.001			7.3 (1)	.007			3.9 (1)	.047
Low	133 (9.5)	1261 (90.5)			1122 (80.5)	272 (19.5)			736 (52.8)	658 (47.2)		
High	72 (15)	407 (85)			412 (86)	67 (14)			278 (58)	201 (42)		
Access to antenatal care information via hotlines or SMS[b]			1.1 (1)	.29			2.5 (1)	.11			0.3 (1)	.59
No	101 (10.2)	887 (89.8)			796 (80.6)	192 (19.4)			529 (53.5)	459 (46.5)		
Yes	104 (11.8)	781 (88.2)			738 (83.4)	147 (16.6)			485 (54.8)	400 (45.2)		
Access to antenatal care information via hospitals' official accounts on social media platforms			18.6 (1)	<.001			33.3 (1)	<.001			13.8 (1)	<.001
	25 (5.5)	432 (94.5)			333 (72.9)	124 (27.1)			213 (46.6)	244 (53.4)		
	180 (12.7)	1236 (87.3)			1201 (84.8)	215 (15.2)			801 (56.6)	615 (43.4)		
Access to antenatal care information from friends or family members			2.1 (1)	.15			0.1 (1)	.81			6.0 (1)	.01
	133 (11.8)	994 (88.2)			925 (82.1)	202 (17.9)			636 (56.4)	491 (43.6)		
Yes	72 (9.7)	674 (90.3)			609 (81.6)	137 (18.4)			378 (50.7)	368 (49.3)		

[a] Percentages reported parenthetically are based on the sum of absolute values of each subcategory (or row) listed under “Characteristic.”

[b] Chi-square test.

[c] Data not available.

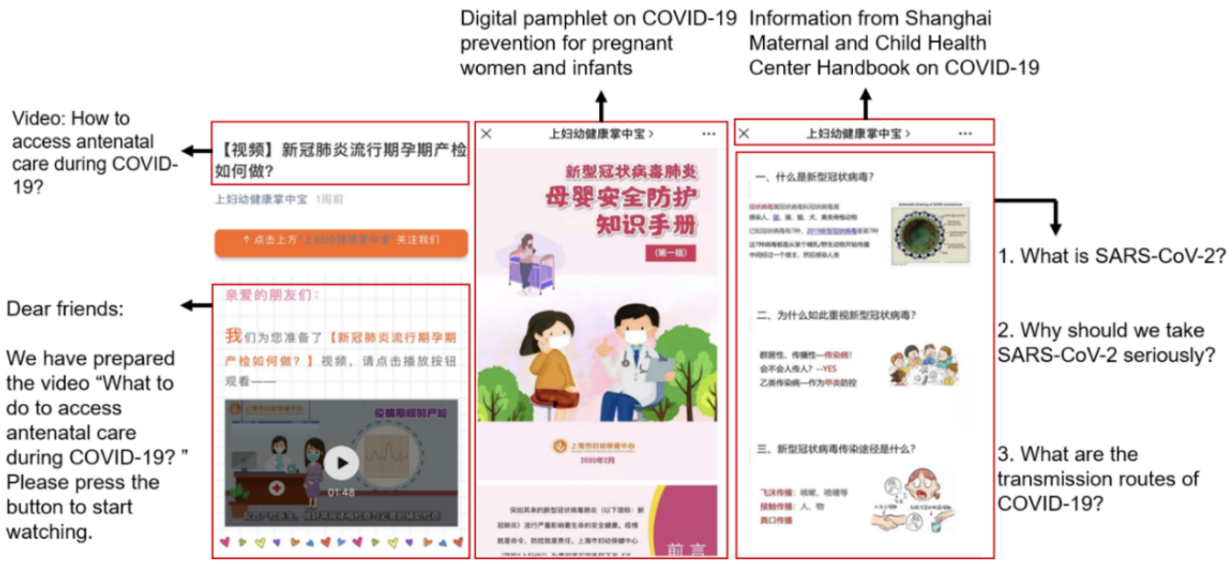
[d] The mean age of 29 years was used to categorize participants into the 2 groups: ≥29 and <29 years.

In all, 75.6% (1416/1873) of the participants reported having ever accessed antenatal health care information via hospitals' WOAs, and 63.2% (1185/1873) of the participants also selected WOAs as the most preferred way to access antenatal care information (Figure 1). Moreover, 47.3% (885/1873) of the participants reported they accessed antenatal care information via hospital hotlines or SMS, and 39.8% (746/1873) accessed this information via friends or family members. Results of the chi-square test showed that women who were relatively older, had received higher education, were employed, in their 1st or 3rd trimester, residing in urban areas, and living in Shanghai area were more likely to access information via hospitals' WOAs. All 3 different approaches of accessing antenatal care information were positively associated with a higher score of COVID-19 self-protection behaviors. Access to information via WOAs was positively associated with a higher score of COVID-19 antenatal care knowledge (Table 1).

Results of the multiple binary logistic regression analyses showed that participants who accessed antenatal care information via hospitals' WOAs were significantly associated with lower risk of perceived stress (odds ratio [OR] 0.46, 95% CI 0.30-0.72; *P*=.001), anxiety (OR 0.53, 95% CI 0.41-0.68; *P*<.001) and depression (OR 0.73, 95% CI 0.59-0.91; *P*=.005; see the table in Multimedia Appendix 2). Furthermore, participants who accessed health care information via hospital hotlines or SMS were found to be significantly associated with a lower risk of anxiety only (OR 0.77, 95% CI 0.60-0.98; *P*=.04;

see the table in Multimedia Appendix 3). Participants who accessed antenatal care information via friends and family members were found to be significantly associated with a higher risk of experiencing depression symptoms (OR 1.32, 95% CI 1.09-1.60; *P*=.004; see Figure 2 and the table in Multimedia Appendix 4).

Figure 1. Examples of antenatal care information provided via a health institution’s official WeChat account. Screenshots of antenatal care information shared via the official WeChat account of Shanghai Maternal and Child Health Center, China, introducing when, where, and how should pregnant women receive antenatal care and COVID-19 prevention strategies for pregnant women and infants.



【视频】新冠肺炎流行期孕期产检如何做?

上妇幼健康掌中宝 1周前

亲爱的朋友们：

我们为您准备了【新冠肺炎流行期孕期产检如何做？】视频，请点击播放按钮观看——

新型冠状病毒肺炎
母婴安全防护
知识手册

一、什么是新型冠状病毒?

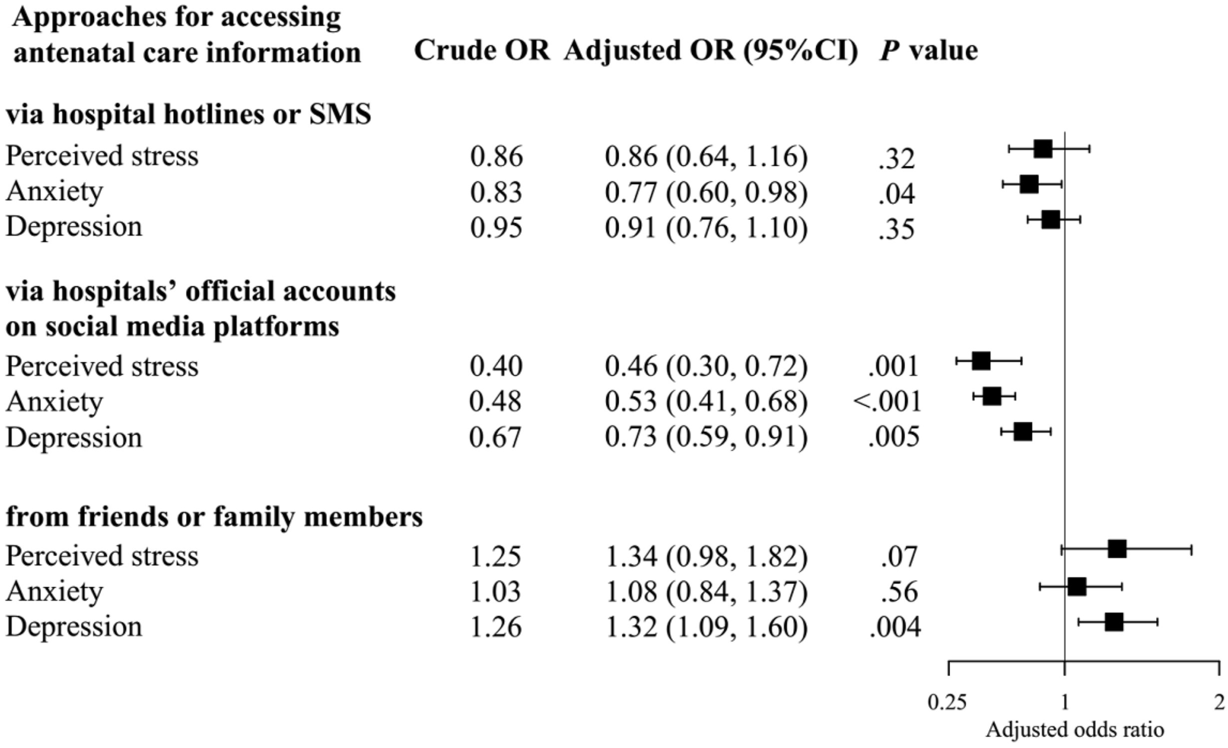
二、为什么如此重视新型冠状病毒?

群居性、传播性—传染病!
会不会人传人? --YES
乙类传染病—作为甲类防控

三、新型冠状病毒传染途径是什么?

飞沫传播：咳嗽、喷嚏等
接触传播：人、物
粪口传播

Figure 2. Approaches used by Chinese pregnant women to access antenatal care information and mental health disorders experienced by them during the COVID-19 epidemic (N=1873). Multiple binary logistic regression analysis controlled for age, education level, employment status, living area, current residence, gestational stage (trimester), parity, pregnancy complications, and scores of COVID-19 self-protection behaviors and antenatal care knowledge. OR: odds ratio.



Approaches for accessing antenatal care information	Crude OR	Adjusted OR (95%CI)	*P* value
via hospital hotlines or SMS			
Perceived stress	0.86	0.86 (0.64, 1.16)	.32
Anxiety	0.83	0.77 (0.60, 0.98)	.04
Depression	0.95	0.91 (0.76, 1.10)	.35
via hospitals’ official accounts on social media platforms			
Perceived stress	0.40	0.46 (0.30, 0.72)	.001
Anxiety	0.48	0.53 (0.41, 0.68)	<.001
Depression	0.67	0.73 (0.59, 0.91)	.005
from friends or family members			
Perceived stress	1.25	1.34 (0.98, 1.82)	.07
Anxiety	1.03	1.08 (0.84, 1.37)	.56
Depression	1.26	1.32 (1.09, 1.60)	.004

Multiple binary logistic regression analysis also showed that factors such as education level of college and above, current residence in Shanghai, and a higher score of COVID-19 antenatal care knowledge were negatively associated with at least one of the mental health disorders assessed in this study, compared with other factors such as a lower education level, current residence outside Shanghai, and a lower score of COVID-19 antenatal care knowledge. Moreover, pregnancy complications were found to be significantly associated with an increased risk of anxiety (Multimedia Appendices 2-4).

Discussion

Principal Findings

Our study provides a snapshot of the mental health status of pregnant women during the early stage of the COVID-19 epidemic in China. The prevalence of perceived stress, anxiety, and depression among the pregnant women was high, reported among 89.1%, 18.1%, and 45.9% of the participants, respectively. In addition, our survey results showed that the most popular way to access antenatal care information was via hospitals' official accounts on social media platforms such as WeChat and Weibo; women who accessed antenatal care information using this approach had a significantly lower risk of perceived stress, anxiety, and depression, after controlling for social economic factors such as age, education, and employment status; obstetric conditions such as gestational stage (trimester) and pregnancy complications; and scores of COVID-19 self-protection behaviors and antenatal care knowledge.

The proportion of women experiencing perceived stress, anxiety, and depression was higher in this study sample, that is, during the COVID-19 epidemic, than that reported in previous Chinese studies in the context of no public health emergency, using the same measurements. For instance, in previous studies, perceived stress was reported at 73.3% [33], anxiety at 11.3% [34,35], and depression at 17.6%-25.4% [32,36]. The higher prevalence of maternal mental health disorders in our study sample was consistent with the findings of other studies conducted during the COVID-19 outbreak [37-40]. The higher risk of maternal mental health disorders might be related to the concerns pregnant women have regarding the COVID-19 pandemic and its impact on the economic and individual social circumstances, such as prolonged stay-at-home, increased exposure to news or rumors, isolation from their social support network and health care facilities, limited access to self-protection resources (eg, face mask and sanitizers), and worry about contracting COVID-19 and mother-to-fetus transmission [16,37]. These concerns and fears could perpetuate stress and anxiety among pregnant women and contribute to increased rates of mental health disorders [38]. Furthermore, compared with the prevalence of depression symptoms (29.6%) in a Chinese study on pregnant women in late January 2020 (ie, at the beginning of the COVID-19 epidemic in China), the prevalence of depression symptoms in our study was much higher (45.9%) [39]. This difference might be due to the time difference between the two studies. Our study was carried out in February 2020, during the peak of the COVID-19 epidemic in China when the number of daily new confirmed cases reached new records and major public health disease control and prevention measures were universally implemented across the country. Although these measures were necessary to contain the spread of the infection, they could also aggravate negative psychological effects among pregnant women [40,41]. There is solid evidence for the link among maternal stress, anxiety, and depression, as well as a number of pregnancy complications and adverse pregnancy and birth outcomes (eg, susceptibility to infection, low birth weight, preterm birth, and impaired cognitive development of the child) [42-44]. The high prevalence of mental health disorders among pregnant women found in this study suggests the needs for timely intervention and preparedness for maternal mental health care in a public health emergency context. Further prospective studies are required to understand the effects of maternal mental health disorders during COVID-19 on pregnancy outcomes and the well-being of the child.

Our findings revealed that access to antenatal care information via hospitals' WOAs was significantly associated with a lower risk of all 3 measured mental health disorders (ie, perceived stress, anxiety, and depression). Obtaining health care information via hospital hotlines or SMS was found to be significantly associated with a lower risk of anxiety only. Nevertheless, both these approaches highlighted the importance of the source of antenatal care information. Hospitals usually are regarded as a reliable and credible source for obtaining health information. The antenatal care knowledge provided by hospitals during the COVID-19 pandemic is likely to enhance the confidence and self-efficacy of pregnant women to deal with the situation. Furthermore, the feeling of being connected with the hospitals and health care providers via hospital WOAs as well as hotline and SMS during the COVID-19 pandemic is likely to provide reassurance and comfort to pregnant women, which may help mitigate distress. However, unlike the hotline, WOAs allow pregnant women to access the information they need at any time as per their convenience. Moreover, compared with SMS, WOAs offer these women the flexibility to obtain more comprehensive information that meets their needs. WOAs can provide substantial information in various ways, including tips, articles, images, and links to videos. These might be some of the possible reasons why access to information via hospitals' WOAs was found to be associated with a higher reduction in mental health disorders among pregnant women compared with access to information via hotlines and SMS.

Surprisingly, we found that pregnant women who obtained antenatal care information via friends and family members were associated with a higher risk of depression. This finding could be partially explained by the fact that women experiencing depression symptoms are more likely to turn to their friends and family for help or that they received more concerns from their friends or family members. Nevertheless, this observation suggested the need for actively disseminating maternal health service information and involving family members in antenatal health education. Being able to access health information from a reliable and credible source has always been a primary service need of new and expecting mothers [45]. The findings of this study highlight the importance of both the trustworthiness of the source and the channels of information acquisition. The findings also suggest that during a public health emergency, reliable information provided by health professional institutions via social media platforms is a feasible and potentially effective way to deliver health care information and services to pregnant women. The findings also signify the importance and necessity of communicating authoritative information by health service providers via social media platforms as one component of epidemic preparedness and response in the future.

The Chinese government and many professional societies have responded swiftly to emergency maternal health interventions during the COVID-19 epidemic. Several professional organizations jointly issued management guidelines for pregnancy during COVID-19 on January 31, 2020 [19]. The National Health Commission of China released a province-wise list of designated hospitals for suspected or confirmed COVID-19 cases among pregnant women on February 12, 2020 [46]. Following these, many MCH institutions have posted service information and guidance for antenatal care on WOAs (Figure 1) [47,48]. However, training and capacity building is urgently needed for translating evidence-based information to health education materials in a timely manner such that it is comprehensible by the public. As an increasing number of health institutions choose to provide telemedicine services such as counselling and consultation, the quality of these remote-based services needs to be closely monitored. Thus, there is a need for more resources to be invested to promote quality antenatal care information dissemination and service provision via web-based platforms.

The findings of this study should be interpreted with caution. This is a web-based survey, wherein participants were self-selected, tended to have received higher education, possessed digital devices, and were possibly more health conscious, all of which might introduce selection bias and not be representative of all pregnant women. The results may also be subjected to recall bias, as the variables are based on self-reporting. The cross-sectional design of this study prevents significant causal associations to be demonstrated. Finally, the residence location of participants was not geographically evenly distributed; only a small proportion of pregnant women resided in Hubei Province—the region most severely affected area by COVID-19 in China.

Conclusions

In the early stage of the COVID-19 epidemic in China, a large proportion of pregnant women experienced high levels of perceived stress, anxiety, and depression. In addition, a significantly lower risk of maternal mental health disorders was found to be associated with accessing professional antenatal care information via hospitals' WOAs. These findings advocate for antenatal health service providers to develop evidence-based health care information and update service provision information on social media platforms in a timely manner. This should also form an essential component of public health preparedness and response to mitigate negative mental health outcomes among pregnant women.

Acknowledgments

This study was funded by the Key Discipline Program of Fifth Round of the Three-Year Public Health Action Plan (Year 2020-2022) of Shanghai (GWV-10.1-XK08). We are grateful to Prof Paul Garner for his immense support in interpreting the results. We are thankful to all participants and health staff involved in this study.

Authors' Contributions

HJ, XQ, and ML conceived the study. HJ, XQ, ML, XX, and LMJ designed the research tools. HJ, LMJ, XNL, WYC, FYY, XWZ, NA, XYL, XYL, ZYX, XLZ, and LLL contributed to the data collection. HJ, ML, XX, XGY, and XNL carried out data analyses and interpretation. HJ, ML, and LMJ drafted the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questions about COVID-19 self-protection behaviors and antenatal care knowledge included in the survey questionnaire.

[DOCX File , 23 KB - jmir_v23i1e18722_app1.docx]

Multimedia Appendix 2

Mental health disorders among Chinese pregnant women accessing antenatal care information via hospitals' official accounts on social media platforms during the COVID-19 outbreak (N=1873).

[DOCX File , 26 KB - jmir_v23i1e18722_app2.docx]

Multimedia Appendix 3

Mental health disorders among Chinese pregnant women accessing antenatal care information via hospitals' hotlines or SMS during the COVID-19 outbreak (N=1873).

[DOCX File , 25 KB - jmir_v23i1e18722_app3.docx]

Multimedia Appendix 4

Mental health disorders among Chinese pregnant women accessing antenatal care information via friends or family members during the COVID-19 outbreak (N=1873).
[DOCX File , 26 KB - jmir_v23i1e18722_app4.docx]

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
2. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. Vital surveillances: The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. China CDC Weekly 2020;2(8):113-122. [doi: 10.46234/ccdcw2020.032]
3. The first-level response to public health emergencies being initiated in 30 provinces of China. News in Chinese. Chinatimes. 2020 Jan 26. URL: http://www.chinatimes.net.cn/article/93775.html [accessed 2021-01-08]
4. Distribution of Novel Coronavirus Pneumonia. Webpage in Chinese. National Health Commission of the Peoples' Republic of China. URL: http://2019ncov.chinacdc.cn/2019-nCoV/6 [accessed 2020-08-30]
5. COVID-19 Strategy Update. World Health Organization. 2020 Apr 14. URL: https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf?sfvrsn=29da3ba0_19 [accessed 2020-08-20]
6. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-09-03]
7. Yuan Z, Xiao Y, Dai Z, Huang J, Zhang Z, Chen Y. Modelling the effects of Wuhan's lockdown during COVID-19, China. Bull. World Health Organ 2020 May 28;98(7):484-494. [doi: 10.2471/blt.20.254045]
8. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
9. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]
10. Zhang W, Yang X, Zhao J, Yang F, Jia Y, Cui C, et al. Depression and psychological-behavioral responses among the general public in China during the early stages of the COVID-19 pandemic: survey study. J Med Internet Res 2020 Sep 04;22(9):e22227 [FREE Full text] [doi: 10.2196/22227] [Medline: 32886066]
11. Favre G, Pomar L, Musso D, Baud D. 2019-nCoV epidemic: what about pregnancies? The Lancet 2020 Feb;395(10224):e40. [doi: 10.1016/s0140-6736(20)30311-1]
12. Qiao J. What are the risks of COVID-19 infection in pregnant women? The Lancet 2020 Mar 07;395(10226):760-762 [FREE Full text] [doi: 10.1016/s0140-6736(20)30365-2]
13. Du L, Gu Y, Cui M, Li W, Wang J, Zhu L, et al. Investigation on demands for antenatal care services among 2 002 pregnant women during the epidemic of COVID-19 in Shanghai (Article in Chinese). Zhonghua Fu Chan Ke Za Zhi 2020 Mar 25;55(3):160-165. [doi: 10.3760/cma.j.cn112141-20200218-00112] [Medline: 32268713]
14. Jamieson DJ, Ellis JE, Jernigan DB, Treadwell TA. Emerging infectious disease outbreaks: old lessons and new challenges for obstetrician-gynecologists. Am J Obstet Gynecol 2006 Jun;194(6):1546-1555 [FREE Full text] [doi: 10.1016/j.ajog.2005.06.062] [Medline: 16731070]
15. Notice to patients of Tongji Hospital. Webpage in Chinese. Tongji Hospital. 2020 Feb 13. URL: https://www.tjh.com.cn/html/2020/0213/29065.shtml#title [accessed 2020-08-25]
16. Samantha K, Dale W, Neil G. Psychological impact of infectious disease outbreaks on pregnant women: Rapid evidence review. medRxiv. Preprint posted online on April 22, 2020 [FREE Full text] [doi: 10.1101/2020.04.16.20068031]
17. Cheung MC. China Social Network Users 2019 - Updates on Douyin, WeChat and Sina Weibo. EMarketer. 2020 Dec 12. URL: https://www.emarketer.com/content/china-social-network-users-2019 [accessed 2020-02-26]
18. Wenjuanxing (Online Survey Platform). URL: https://www.wjx.cn/app/survey.aspx [accessed 2021-01-13]
19. MaternalFetal Experts Committee, Chinese Medical Association ObstetricsGynecology Society, Chinese Medical Association Perinatal Medicine Society, Chinese Medical Association Editorial Board of Chinese Journal of Perinatal Medicine. Proposed management of 2019-novel coronavirus infection during pregnancy and puerperium (Article in Chinese). Chin J Perinat Med 2020 Jan;23(2):73-79. [doi: 10.3760/cma.j.issn.1007-9408.2020.02.001]
20. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
21. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
22. Yang TZ, Zheng JZ. Health education theory and approach. 1st edition. Book in Chinese. Hangzhou: Zhejiang University Press; 2004:200-203.
23. Yang TZ, Huang HT. An epidemiological study on stress among urban residents in social transition period (Article in Chinese). Zhonghua Liu Xing Bing Xue Za Zhi 2003 Sep;24(9):760-764. [Medline: 14521764]

24. Nengjiang L, Yuchnan L, Haolan Z, Yajun X. Survey analysis and solutions towards the current state of medical personnel's work stress (Article in Chinese). J Med Thero & Prac 2014;27(16):2126-2128 [FREE Full text]
25. Wang Z, Chi Y. Self-rating anxiety scale (SAS) (Article in Chinese). Shanghai Archives of Psychiatry 1984:73-74.
26. Zung WW. A rating instrument for anxiety disorders. Psychosomatics 1971 Nov;12(6):371-379. [doi: 10.1016/S0033-3182(71)71479-0]
27. Olatunji BO, Deacon BJ, Abramowitz JS, Tolin DF. Dimensionality of somatic complaints: factor structure and psychometric properties of the Self-Rating Anxiety Scale. J Anxiety Disord 2006 Jan;20(5):543-561. [doi: 10.1016/j.janxdis.2005.08.002] [Medline: 16198532]
28. Zhao Y, Li M, Chen Y, Wei Y, Feng Q. The pre-labor status of anxiety and depression in pregnant women and their influencing factors (Article in Chinese). Journal of Practical Obstetrics and Gynecology 2006(10):608-610 [FREE Full text]
29. Lee DTS, Yip SK, Chiu HFK, Leung TYS, Chan KPM, Chau IOL, et al. Detecting postnatal depression in Chinese women. Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. Br J Psychiatry 1998 May;172:433-437. [doi: 10.1192/bjp.172.5.433] [Medline: 9747407]
30. Cox J. Validation of the Edinburgh postnatal depression scale (EPDS) in non-postnatal women. Journal of Affective Disorders 1996 Jul 29;39(3):185-189. [doi: 10.1016/0165-0327(96)00008-0]
31. Xiong, MD, DrPH X, Harville, PhD EW, Mattison, MD DR, Elkind-Hirsch, PhD K, Pridjian, MD G, Buekens, MD, PhD P. Hurricane Katrina experience and the risk of post-traumatic stress disorder and depression among pregnant women. Am J Disaster Med 2010 May 01;5(3):181-187. [doi: 10.5055/ajdm.2010.0022]
32. Yang T, He H, Ma CY, Ji C, Zeng S. Risk factors of prenatal depression and anxiety in pregnant women (Article in Chinese). Chinese Journal of Mental Health 2015;29:246-250 [FREE Full text]
33. Chen MQ, Ma XY, Xiong X, Peng JX, Zhu XL, Han JJ. Study on related factors of depression during early pregnancy (Article in Chinese). Maternal and Child Health Care of China 2017;03:451-454.
34. Ma X, Wang Y, Hu H, Tao XG, Zhang Y, Shi H. The impact of resilience on prenatal anxiety and depression among pregnant women in Shanghai. J Affect Disord 2019 May 01;250:57-64. [doi: 10.1016/j.jad.2019.02.058] [Medline: 30831542]
35. Ma J, Wei C, Zhu D, Sun X, Li Q, Yang F. An investigation and analysis of prevalence of anxiety and depression and their influencing factors among pregnant women in early, middle and late pregnancy (Article in Chinese). Chinese Journal of Women and Child Health Research 2010;21(4):432-434.
36. Zhao Y, Kane I, Wang J, Shen B, Luo J, Shi S. Combined use of the postpartum depression screening scale (PDSS) and Edinburgh postnatal depression scale (EPDS) to identify antenatal depression among Chinese pregnant women with obstetric complications. Psychiatry Res 2015 Mar 30;226(1):113-119. [doi: 10.1016/j.psychres.2014.12.016] [Medline: 25677395]
37. Lebel C, MacKinnon A, Bagshawe M, Tomfohr-Madsen L, Giesbrecht G. Elevated depression and anxiety symptoms among pregnant individuals during the COVID-19 pandemic. J Affect Disord 2020 Dec 01;277:5-13 [FREE Full text] [doi: 10.1016/j.jad.2020.07.126] [Medline: 32777604]
38. Saccone G, Florio A, Aiello F, Venturella R, De Angelis MC, Locci M, et al. Psychological impact of coronavirus disease 2019 in pregnant women. Am J Obstet Gynecol 2020 Aug;223(2):293-295 [FREE Full text] [doi: 10.1016/j.ajog.2020.05.003] [Medline: 32387321]
39. Wu Y, Zhang C, Liu H, Duan C, Li C, Fan J, et al. Perinatal depressive and anxiety symptoms of pregnant women during the coronavirus disease 2019 outbreak in China. Am J Obstet Gynecol 2020 Aug;223(2):240.e1-240.e9 [FREE Full text] [doi: 10.1016/j.ajog.2020.05.009] [Medline: 32437665]
40. Corbett GA, Milne SJ, Hehir MP, Lindow SW, O'connell MP. Health anxiety and behavioural changes of pregnant women during the COVID-19 pandemic. Eur J Obstet Gynecol Reprod Biol 2020 Jun;249:96-97 [FREE Full text] [doi: 10.1016/j.ejogrb.2020.04.022] [Medline: 32317197]
41. Thapa SB, Mainali A, Schwank SE, Acharya G. Maternal mental health in the time of the COVID-19 pandemic. Acta Obstet Gynecol Scand 2020 Jul;99(7):817-818 [FREE Full text] [doi: 10.1111/aogs.13894] [Medline: 32374420]
42. McEwen BS. Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. Eur J Pharmacol 2008 Apr 07;583(2-3):174-185 [FREE Full text] [doi: 10.1016/j.ejphar.2007.11.071] [Medline: 18282566]
43. Glover V. Maternal depression, anxiety and stress during pregnancy and child outcome; what needs to be done. Best Pract Res Clin Obstet Gynaecol 2014 Jan;28(1):25-35. [doi: 10.1016/j.bpobgyn.2013.08.017] [Medline: 24090740]
44. Madigan S, Oatley H, Racine N, Fearon RP, Schumacher L, Akbari E, et al. A meta-analysis of maternal prenatal depression and anxiety on child socioemotional development. J Am Acad Child Adolesc Psychiatry 2018 Sep;57(9):645-657.e8. [doi: 10.1016/j.jaac.2018.06.012] [Medline: 30196868]
45. Jiang H, Li M, Wen LM, Baur LA, He G, Ma X, et al. A short message service intervention for improving infant feeding practices in Shanghai, China: planning, implementation, and process evaluation. JMIR Mhealth Uhealth 2018 Oct 29;6(10):e11039 [FREE Full text] [doi: 10.2196/11039] [Medline: 30373728]
46. All provinces (autonomous regions, municipalities) designated hospitals for maternity examinations and inpatient delivery for suspected or confirmed new coronavirus infections. National Health Commission of People's Republic of China. 2020 Feb 11. URL: http://www.nhc.gov.cn/fys/s7901/202002/d4e8fae0649745118f63136660f4c3b0.shtml [accessed 2020-02-26]

47. Guidelines for preventing COVID-19 for pregnant women. Webpage in Chinese. WeChat official account of Chinese Center for Disease Control and Prevention. 2020 Feb 14. URL: https://mp.weixin.qq.com/s/EkCSVJjDjoqAbF-92U1KIw [accessed 2020-02-29]
48. Prevention measures for expectant mothers to fight against COVID-19. WeChat official account of Maternal and Child Health Hospital of Minhang District, Shanghai. 2020 Feb 09. URL: https://mp.weixin.qq.com/s/9niZ8WzG0OOh2XuQwMyQWA [accessed 2020-02-29]

Abbreviations

CPSS: Chinese Perceived Stress Scale
EDS: Edinburgh Depression Scale
MCH: maternal and child health
OR: odds ratio
QR: quick response
SAS: Self-Rating Anxiety Scale (SAS)
WOA: WeChat or Weibo official account

Edited by R Kukafka; submitted 12.06.20; peer-reviewed by C Son, Y Bao; comments to author 19.08.20; revised version received 06.09.20; accepted 17.12.20; published 18.01.21.

Please cite as:
Jiang H, Jin L, Qian X, Xiong X, La X, Chen W, Yang X, Yang F, Zhang X, Abudukelimu N, Li X, Xie Z, Zhu X, Zhang X, Zhang L, Wang L, Li L, Li M
Maternal Mental Health Status and Approaches for Accessing Antenatal Care Information During the COVID-19 Epidemic in China: Cross-Sectional Study
J Med Internet Res 2021;23(1):e18722
URL: http://www.jmir.org/2021/1/e18722/
doi:10.2196/18722
PMID:33347423

Original Paper

Mental Health Among Medical Professionals During the COVID-19 Pandemic in Eight European Countries: Cross-sectional Survey Study

Svenja Hummel[1*]; Neele Oetjen[1*], MSc; Junfeng Du[1,2], MSc; Elisabetta Posenato[1], MSc; Rosa Maria Resende de Almeida[3], MSc; Raquel Losada[3], MSc; Oscar Ribeiro[4], PhD; Vincenza Frisardi[5], MA, MD; Louise Hopper[6], PhD; Asarnusch Rashid[7], PhD; Habib Nasser[8], PhD; Alexandra König[9], PhD; Gottfried Rudofsky[10], MD; Steffi Weidt[11], Dr med, PD; Ali Zafar[1], MSc, MA; Nadine Gronewold[1], MSc; Gwendolyn Mayer[1], Dipl-Psych; Jobst-Hendrik Schultz[1], Dr med, PD

[1]Department of General Internal and Psychosomatic Medicine, Heidelberg University Hospital, Heidelberg, Germany
[2]Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[3]Intras Foundation, Valladolid, Spain
[4]Center for Health Technology and Services Research, Department of Education and Psychology, University of Aveiro, Aveiro, Portugal
[5]Geriatric and NeuroRehabilitation Department, AUSL-IRCCS Reggio Emilia, Emilia, Italy
[6]School of Psychology, Faculty of Science and Health, Dublin City University, Dublin, Ireland
[7]ZTM Bad Kissingen GmbH, Bad Kissingen, Germany
[8]R&D Department, RDIUP, Les Mureaux, France
[9]Memory Clinic, Institut Claude Pompidou, Nice, France
[10]Clinic for endocrinology and metabolic disorders, Kantonspital Olten, Olten, Switzerland
[11]Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital Zurich, University of Zurich, Zurich, Switzerland
[*]these authors contributed equally

Corresponding Author:
Neele Oetjen, MSc
Department of General Internal and Psychosomatic Medicine
Heidelberg University Hospital
Im Neuenheimer Feld 410
Heidelberg, 69120
Germany
Phone: 49 62215632408
Email: Neele.Oetjen@med.uni-heidelberg.de

Abstract

Background: The death toll of COVID-19 topped 170,000 in Europe by the end of May 2020. COVID-19 has caused an immense psychological burden on the population, especially among doctors and nurses who are faced with high infection risks and increased workload.

Objective: The aim of this study was to compare the mental health of medical professionals with nonmedical professionals in different European countries during the COVID-19 pandemic. We hypothesized that medical professionals, particularly those exposed to COVID-19 at work, would have higher levels of depression, anxiety, and stress. We also aimed to determine their main stressors and most frequently used coping strategies during the crisis.

Methods: A cross-sectional online survey was conducted during peak COVID-19 months in 8 European countries. The questionnaire included demographic data and inquired whether the participants were exposed to COVID-19 at work or not. Mental health was assessed via the Depression Anxiety Stress Scales32 (23.53)–21 (DASS-21). A 12-item checklist on preferred coping strategies and another 23-item questionnaire on major stressors were completed by medical professionals.

Results: The sample (N=609) consisted of 189 doctors, 165 nurses, and 255 nonmedical professionals. Participants from France and the United Kingdom reported experiencing severe/extremely severe depression, anxiety, and stress more often compared to those from the other countries. Nonmedical professionals had significantly higher scores for depression and anxiety. Among medical professionals, no significant link was reported between direct contact with patients with COVID-19 at work and anxiety, depression, or stress. “Uncertainty about when the epidemic will be under control” caused the most amount of stress for health care professionals while “taking protective measures” was the most frequently used coping strategy among all participants.

Conclusions: COVID-19 poses a major challenge to the mental health of working professionals as a considerable proportion of our participants showed high values for depression, anxiety, and stress. Even though medical professionals exhibited less mental stress than nonmedical professionals, sufficient help should be offered to all occupational groups with an emphasis on effective coping strategies.

(J Med Internet Res 2021;23(1):e24983) doi:10.2196/24983

KEYWORDS

mental health; COVID-19; Europe; medical professionals; stress; depression; anxiety; coping; stressors

Introduction

Background

The death toll of COVID-19 topped 170,000 in Europe by the end of May 2020 [1]. Italy, Spain, France and the United Kingdom were among the worst affected countries with respect to high infection rates and overburdened health care systems [2-4]. Deaths of approximately 0.9 per 1 million inhabitants for Spain and France, 1.45 for Italy, and 3.94 for the United Kingdom were reported at the end of May. The death rate in Portugal was high as well with 1.32 deaths per million inhabitants [5]. Nevertheless, it was not one of the most affected countries [6,7]. Mortality in Germany was 0.47 deaths per 1 million inhabitants, Austria 0.52, and Switzerland 0.3 [5]; these values are lower but nevertheless alarming. As the number of cases soared, national governments introduced widespread restrictions to control the virus spread such as closing down borders, social distancing, travel restrictions, wearing of masks, working from home, and closure of public facilities [8,9]. The impending risk of infection, increasing number of COVID-19 cases, and the overburdening of health care systems created an unprecedented situation, which impacted not only everyday life but also the psychological welfare of the general population.

COVID-19 and Mental Health

A recent study in 41 countries showed high stress levels in the general population during the COVID-19 pandemic, similar to those reported during the severe acute respiratory syndrome (SARS) epidemic in 2003 [10]. Depressive symptoms increased from 8.5% before COVID-19 to 27.8% during COVID-19 in a US study [11]. Studies have shown a similar picture for Europe. An Italian study concluded that leaving home for work was associated with increased stress, which may be due to the fear of getting infected [12]. In a study from Spain during the lockdown, about 34% and 21% of the participants reported moderate to extremely severe depression and anxiety, respectively [13]. A study from the United Kingdom reported high mental distress during the lockdown [14]. The first part of a French study in March 2020 found that the prevalence of anxiety among the general population was twice as high as reported in a study before COVID-19 [15]. In comparison, a study in Hong Kong found that about one-third of SARS survivors still suffered from moderate to severe or severe anxiety and/or depressive symptoms 4 weeks after their recovery [16]. Even 30 months after exposure to SARS, psychiatric disorders were prevalent among SARS survivors [17].

COVID-19 and Medical Professionals

“Breathe. It’s what we all want these days, doctors and patients, nurses and care workers. All of us. We want air,” wrote an Italian frontline doctor from Milan in April during the peak of COVID-19 while sharing his routine of wearing a mask all the time [18].

Among the health care workers struggling to cope with the situation of doing their jobs while trying to protect themselves and their families [19], one would suspect doctors and nurses to be the most affected psychologically. During SARS, physicians who had direct contact with infected patients expressed greater mental distress, more stigmatization, and more worries about infecting their family [20]. This was confirmed during the COVID-19 pandemic in Italy where frontline health care workers reported posttraumatic stress symptoms [21]. Thousands of medical professionals were sent into quarantine after contracting COVID-19 in Italy, France, and Spain [22]. The fact that they came so close to the disease put their mental health at a higher risk than the general population [19]. The psychological challenges of the pandemic have been described for nurses and doctors in Europe [23,24]. Stronger effects on the mental health of medical health workers compared to nonmedical health workers [25] and a high prevalence of mental health symptoms among physicians and other medical staff have also been reported in China [26,27].

Coping Strategies and Major Stressors

A 2012 study with nurses showed that negative coping strategies led to higher levels of mental distress whereas positive coping was partly negatively correlated with depression and anxiety levels [28]. During SARS, physicians in Toronto were mainly concerned about disparities in the health care of non-SARS patients because of the special situation [20] while nurses in Taiwan experienced “endangering of colleagues” as a major stressor [29]. Getting information about the virus and sticking to infection control procedures seemed to be important coping strategies [20,29]. During the COVID-19 pandemic, coping strategies and stressors of health care workers have been investigated in China, showing that health care workers worried a lot about possibly infecting their families and were highly stressed by witnessing the deaths of infected patients. Adhering to protective measures and learning more about COVID-19 were most often used to cope [30,31].

Objectives

At a time when public health systems are overburdened in the fight against COVID-19 [2], physically and mentally healthy professionals are essential for the provision of reliable and efficient health care services. Physician burnout has been linked with medical errors [32] and further harmful effects for coworkers, patients, and the whole health care system [33]. The influence of COVID-19 on mental health in individual European countries and/or in individual population groups has been assessed, but a clear answer to the question of an overall impression of mental health in Europe during the peak months of the pandemic is still lacking. As global rates of infections rise once again and an effective vaccine remains unavailable, this question gains further relevance for determining mental health care needs for working professionals in the near future.

This study aimed to explore medical and nonmedical professionals' mental health in different European countries during the 3-month state of emergency due to the pandemic and whether or not it was influenced by exposure to the virus at work. We hypothesized that medical professionals, particularly those exposed to COVID-19 at work, would show higher scores in depression, anxiety, and stress compared to nonmedical professionals. Moreover, we investigated which aspects of the COVID-19 pandemic worried medical professionals the most while at work and which coping strategies they most frequently employed. By uncovering these stressors and coping strategies, it might be possible to devise policies to prepare and support medical professionals better for future crises.

Methods

Study Design

We used a cross-sectional, multilingual survey design to investigate the mental health of working professionals in 8 European countries (Germany, the United Kingdom, Spain, France, Portugal, Austria, Italy, and Switzerland) during 3 months of the COVID-19 crisis between April 1 and June 20, 2020. The focus was on medical professionals and whether they were exposed to COVID-19 at work. Additionally, we asked about the most stressful aspects of work and coping strategies most often used. Ethical approval for this study was granted by the Ethics Commission of the Medical Faculty of Heidelberg University (S-361/2020).

Participants

The participants were recruited online via public social networking groups and via the authors' European contacts with partner organizations from international joint projects.

The sample (N=609) included 354 people with medical professions, including 189 doctors and 165 nurses (including geriatric care), and 255 people with nonmedical occupations (eg, teachers, office staff, psychologists, retired persons, social workers). Participants were aged 18-84 years (median 41 years) with 151 males and 458 females.

The percentage distribution of participants and professional groups in different European countries is summarized in Table 1.

Table 1. Distribution of professional groups within European countries.

Country	Medical professionals, n (%)	Nonmedical professionals, n (%)	Total, n
Germany	100 (73.53)	36 (26.47)	136
Austria	6 (28.57)	15 (71.43)	21
Switzerland	33 (82.50)	7 (17.50)	40
France	15 (28.85)	37 (71.15)	52
Italy	142 (89.31)	17 (10.69)	159
Spain	28 (28.28)	71 (71.72)	99
Portugal	25 (54.35)	21 (45.65)	46
United Kingdom	5 (8.93)	51 (91.07)	56

Measurements

The survey consisted of a questionnaire derived from several validated instruments, with added items on demographics (eg, gender, age, marital status, etc) and a question on whether or not the participants were exposed to patients with COVID-19 at work.

Mental stress was assessed via the Depression Anxiety Stress Scales–21 (DASS-21)—a shorter version of the DASS-42 [34]—which is available in different languages. The DASS-21 consists of 21 items, which can be divided into 3 subscales, each containing 7 items to measure depression (eg, "I couldn't seem to experience any positive feeling at all"), anxiety (eg, "I was aware of dryness of my mouth"), and stress (eg, "I found it difficult to relax"). The responses are rated on a 4-point Likert scale ranging from 0 (did not apply to me at all) to 3 (applied to me very much or most of the time) to indicate how much the statement applied to the participant over the past week. According to Lovibond and Lovibond [34], the authors of the questionnaire, scores for the subscales are calculated by adding the answers of the 7 items for each subscale and then multiplying the result by 2 to get the total score for each participant for comparison to the DASS normative data [35].

To determine the most important stressors for medical staff, we used a questionnaire similar to the one used in a study by Lee et al [29] on SARS, which contains specific items for medical staff. The questionnaire consisted of 23 items. On a Likert scale from 0 (not at all) to 4 (very much), the participants indicated how often they thought about or were concerned about the individual stressors in their everyday life or at work. Given the focus of the study, this questionnaire was completed only by the medical staff participants.

Also based on Lee et al [29], we derived a questionnaire on coping strategies where participants could respond to 12 items using a scale from 0 (almost never) to 3 (almost always). Although all participants completed the coping strategies questionnaires, only results from the medical staff participants are presented here considering the aims and objectives of the study.

Procedure

All questionnaires were translated from English by native speakers or professional translators for use in the respective countries. The survey was made available online via the Soscisurvey.de [36] platform. Consent to participate was obtained online. The English and German questionnaires were distributed at the beginning of April 2020, followed shortly after by the Italian version. The surveys in Spain and Portugal were launched in mid-April and in France in mid-May.

Data Analysis

Using the Lovibond and Lovibond [34] method, the depression, anxiety, and stress subscales on DASS-21 were, according to individual sum scores, categorized as normal, mild, moderate, severe, and extremely severe [34]. These subscales were then grouped as normal/mild, moderate, and severe/very severe.

We created two groups: medical professionals consisting of doctors and nurses; and nonmedical professionals, which included other jobs in health care, volunteers, nonmedical staff, and community health care workers.

Descriptive analysis (including means, SDs, and frequencies) and inference statistics (multivariate analysis of variance [MANOVA]) were calculated using SPSS, version 24 (IBM Corp) [37].

Results

Distress Levels Across Surveyed European Countries

Across all surveyed countries, 65% (n=396) of the participants reported a normal/mild level of depression, followed by 18% (n=108) with moderate and 17% (n=105) with severe/extremely severe depression. Regarding anxiety, 63% (n=386) reported a normal/mild level of anxiety, 15% (n=91) a moderate level, and 22% (n=132) a severe/extremely severe level. In terms of stress, 59% (n=356) reported a normal/mild level, 14% (n=87) a moderate level, and 27% (n=166) a severe/extremely severe level. Tables 2-4 shows the mean scores for depression, anxiety, and stress for each of the 8 European countries as well as the percentage of participants assigned to the groups normal/mild, moderate, and severe/extremely severe.

Table 2. Depression levels in different European countries assessed using the Depression Anxiety Stress Scales–21.

Country	Mean (SD)[a]	Normal/mild, n (%)	Moderate, n (%)	Severe/very severe, n (%)
Germany	11.49 (8.91)	87 (63.97)	28 (20.59)	21 (15.44)
Austria	7.33 (8.23)	17 (80.95)	3 (14.29)	1 (4.76)
Switzerland	7.45 (8.68)	31 (77.50)	6 (15.00)	3 (7.50)
France	17.42 (11.63)	22 (42.31)	11 (21.15)	19 (36.54)
Italy	10.03 (9.30)	119 (74.84)	19 (11.95)	21 (13.21)
Spain	8.51 (8.98)	75 (75.76)	14 (14.14)	10 (10.10)
Portugal	12.26 (8.46)	24 (52.17)	14 (30.44)	8 (17.39)
United Kingdom	17.64 (11.04)	21 (37.50)	13 (23.21)	22 (39.29)
Total	11.34 (9.90)	396 (65.02)	108 (17.73)	105 (17.24)

Table 3. Anxiety levels in different European countries assessed using the Depression Anxiety Stress Scales–21.

Country	Mean (SD)	Normal/mild, n (%)	Moderate, n (%)	Severe/very severe, n (%)
Germany	8.44 (7.94)	85 (62.50)	19 (13.97)	32 (23.53)
Austria	4.86 (5.68)	15 (71.43)	5 (23.81)	1 (4.76)
Switzerland	4.10 (6.13)	34 (85.00)	3 (7.50)	3 (7.50)
France	11.39 (10.53)	27 (51.92)	10 (19.23)	15 (28.85)
Italy	7.64 (8.39)	110 (69.18)	25 (15.72)	24 (15.09)
Spain	10.04 (10.54)	60 (60.61)	12 (12.12)	27 (27.27)
Portugal	9.83 (8.59)	23 (50.00)	11 (23.91)	12 (26.09)
United Kingdom	10.36 (9.69)	32 (57.14)	6 (10.71)	18 (32.14)
Total	8.61 (9.00)	386 (63.38)	91 (14.94)	132 (21.67)

Table 4. Stress levels in different European countries assessed using the Depression Anxiety Stress Scales–21.

Country	Mean (SD)	Normal/mild, n (%)	Moderate, n (%)	Severe/very severe, n (%)
Germany	17.13 (9.94)	76 (55.88)	27 (19.85)	33 (24.27)
Austria	14.10 (7.96)	17 (80.95)	1 (4.76)	3 (14.29)
Switzerland	11.40 (11.29)	32 (80.00)	2 (5.00)	6 (15.00)
France	21.77 (12.24)	25 (48.08)	4 (7.69)	23 (44.23)
Italy	17.25 (10.46)	91 (57.23)	27 (16.98)	41 (25.79)
Spain	16.42 (10.45)	62 (62.63)	13 (13.13)	24 (24.24)
Portugal	20.78 (10.95)	24 (52.17)	3 (6.52)	19 (41.30)
United Kingdom	18.86 (10.13)	29 (51.79)	10 (17.86)	17 (30.36)
Total	17.40 (10.71)	356 (58.46)	87 (14.29)	166 (27.26)

Comparison of Medical With Nonmedical Professionals

A one-way MANOVA showed a significant main effect for profession ($F_{3,605}$=5.019, P=.002, Wilk's Λ=0.976). The effects were significant for depression ($F_{1,607}$=7.929, P=.005) and anxiety ($F_{1,607}$=5.87, P=.02], which indicated that medical professionals were less depressed (mean 10.39, SD 9.12) compared to nonmedical staff (mean 12.67, SD 10.77), as well as less anxious (mean 7.90, SD 8.36) than nonmedical staff (mean 9.65, SD 9.66). No statistically significant differences were found between medical professionals who had or had no exposure to COVID-19 at work ($F_{3,350}$=0.525, P=.67, Wilk's Λ=0.996).

Table 5 shows the 3 subscales of the DASS-21 and the percentage of medical and nonmedical professionals categorized as normal/mild, moderate, severe/extremely severe for each of these subscales.

Table 5. Overview of depression, anxiety, and stress levels for medical (n=345) and nonmedical professionals (n=255) assessed using the Depression Anxiety Stress Scales–21.

Participants	Mean (SD)	Normal/mild, n (%)	Moderate, n (%)	Severe/very severe, n (%)
Depression				
Medical professionals	10.39 (9.12)	246 (69.49)	60 (16.95)	48 (13.56)
Nonmedical professionals	12.67 (10.77)	150 (58.82)	48 (18.82)	57 (22.35)
Total	11.34 (9.90)	396 (65.03)	108 (17.73)	105 (17.24)
Anxiety				
Medical professionals	7.90 (8.36)	240 (67.80)	49 (13.84)	65 (18.36)
Nonmedical professionals	9.65 (9.66)	146 (57.26)	42 (16.47)	67 (26.28)
Total	8.61 (9.00)	386 (63.38)	91 (14.94)	132 (21.68)
Stress				
Medical professionals	17.10 (10.51)	208 (58.76)	55 (15.54)	91 (25.71)
Nonmedical professionals	17.80 (10.98)	148 (58.04)	32 (12.55)	75 (29.41)
Total	17.40 (10.71)	356 (58.46)	87 (14.29)	166 (27.26)

Stress Factors for Medical Professionals

The highest rated stressors were "uncertainty about when the epidemic will be under control" (mean 2.27, SD 0.85), "worry about inflicting COVID-19 on family" (mean 2.25, SD 0.99), "worry about nosocomial spread" (mean 2.02, SD 0.92) and a "frequent modification of infection control procedures" (mean 2.02, SD 0.89). Participants were least concerned about themselves (mean 1.12, SD 1.04) or coworkers (mean 1.25, SD 0.97) showing COVID-19–like symptoms, conflicts at work as the equivocal definition of responsibility between doctors and nurses (mean 1.19, SD 1.04), and blame from their commanding officers (mean 0.70, SD 0.95). An overview of all stressors in the order of reported severity can be found in Table 6.

Table 6. Stressors for doctors and nurses during COVID-19.

Items[a]	Responses, n	Mean (SD)
Uncertainty about when the epidemic will be under control	350	2.27 (0.85)
Worry about inflicting COVID-19 on family	351	2.25 (0.99)
Worry about nosocomial spread	348	2.02 (0.92)
Frequent modification of infection control procedures	350	2.02 (0.89)
Protective gears cause physical discomfort	349	1.75 (1.02)
Deterioration of patients' condition	347	1.70 (1.00)
Worry about lack of proper knowledge and equipment	349	1.67 (1.04)
Worry about being negligent and endangering patients	350	1.66 (1.07)
Worry about getting infected	349	1.62 (1.03)
Patients' emotional reaction	348	1.57 (0.96)
Worry about lack of manpower	348	1.56 (1.05)
Unclear documentation and reporting procedures	347	1.54 (1.01)
Patient families' emotional reaction	346	1.52 (1.01)
Coworkers being emotionally unstable	348	1.52 (0.97)
Being without a properly fitted environment	348	1.51 (1.08)
Conflict between duty and safety	348	1.49 (1.07)
Worry about being negligent and endangering coworkers	351	1.48 (1.03)
Be infected by colleagues	349	1.31 (1.02)
Protective gear being a hindrance to providing quality care	349	1.28 (1.05)
Coworkers displaying COVID-19–like symptoms	347	1.25 (0.97)
Equivocal definition of the responsibility between doctors and nurses	346	1.19 (1.04)
Yourself displaying COVID-19–like symptoms	347	1.12 (1.04)
Blame from commanding officers	345	0.70 (0.95)

[a] Responses to the question: "When you think about COVID-19 in your life and work, how often did you think or worry about the following things?" (0=not at all, 3=very much).

Coping Strategies of Medical Professionals

The most frequently used strategies were "taking protective measures" (mean 2.70, SD 0.57) and "actively acquiring more knowledge about COVID-19" (mean 2.34, SD 0.80). Alcohol and drugs were the least used strategy (mean 0.32, SD 0.60). Table 7 summarizes these results.

Table 7. Doctors' and nurses' coping strategies during COVID-19 (n=354).

Items[a]	Mean (SD)
Taking protective measures (washing hands, wearing a mask, taking own temperature, etc)	2.7 (0.57)
Actively acquiring more knowledge about COVID-19 (symptoms, transmission pathway, etc)	2.34 (0.80)
Video-chatting with family and friends by phone to share concerns and support	1.84 (0.87)
Engaging in recreational activities (online shopping, social media, internet surfing, etc)	1.62 (0.94)
Engaging in health-promoting behaviors (more rest, exercise, balanced diet, etc)	1.55 (0.99)
Switching thoughts and facing the situations with a positive attitude	1.54 (0.89)
Limiting oneself from watching too much news about COVID-19	1.37 (0.96)
Distracting oneself from thinking about COVID-19 issues by suppression or keeping busy	1.30 (0.92)
Acquiring mental health knowledge and information	1.01 (0.95)
Venting emotions by crying, screaming, smashing things, etc	0.50 (0.81)
Practicing relaxation methods (meditation, yoga, tai chi, etc)	0.46 (0.82)
Using alcohol or drugs	0.32 (0.60)

[a] Responses to the question: "When you think about COVID-19 in your life and work, how often did you use or try to use the following methods to handle the situation?" (1=almost never, 4=almost always).

Discussion

Overview

This study focused on doctors and nurses who were and are facing exceptional physical and mental challenges during the COVID-19 pandemic. In order to gain a deeper understanding of their situation, we investigated the perceived burden of different stressors on medical professionals as well as their coping strategies. Additionally, a general overview of the experienced stress, anxiety, and depression in various European countries during peak months of the pandemic was presented.

Mental Health of Medical Professionals

The majority of doctors and nurses reported a normal to mild level of psychological strain, but about one-third expressed a moderate to extremely severe level of distress. Mental distress is associated with patient safety and a higher probability of medical errors [32,33]. Considering that long-term effects such as posttraumatic stress disorders are not uncommon among this professional group [38] and that the COVID-19 infection rate may increase again, our results should be taken seriously.

However, surprisingly, the mean scores for depression and anxiety among health care professionals were significantly lower than among nonmedical professionals. Regarding the level of stress, there was no significant difference between the two groups. These results are encouraging in the sense that the medical professionals—although confronted with difficult challenges and risks [19,39,40]—seemed to be mentally well prepared to handle the pandemic situation. A possible explanation could be, however, that their medical background helped them to better understand and classify COVID-19–related information when compared to their nonmedical counterparts. When they could feel self-sufficient, the situation appeared more manageable for them. A study with SARS survivors concluded that a better sense of self-care and self-efficacy led to better psychological adjustment to the situation [41]. In accordance with our results, in a study from Singapore, nonmedical health care professionals had a higher prevalence of anxiety than medical health care professionals during COVID-19. The authors believed that nonmedical health care professionals might have had less access to psychological support, less direct information about the situation and received less training on personal protective equipment and infection control measures [42]. In addition, previous European studies show that there is also an increased psychological burden during COVID-19 in the general population. For example, female gender and younger age were identified as risk factors [43,44], which were represented in large numbers in our study population. It has therefore already been recommended to take care of the mental health of the general population as well as special population groups [44].

Interestingly, we did not find any significant association between direct contact with COVID-19–infected patients at work and scores for anxiety, depression, or stress among medical professionals despite a previous study with health workers during the pandemic reporting more psychological distress when there was direct exposure to infected patients [21]. Not only in medical departments specializing in COVID-19 but in all medical units, protective measures such as permanent wearing of face masks, bans on visits to hospitals and nursing homes, and stricter hygiene regulations were made obligatory for medical personnel in the surveyed countries. In addition, because of the considerable number of deaths of doctors and nurses [45], COVID-19 would have been perceived by medical professionals as a kind of ever-present threat and not only when in direct contact with infected patients.

Major Stressors and Coping Strategies

"Uncertainty about when the epidemic will be under control" and "worry about inflicting COVID-19 on family" were at the top of the list when medical professionals were asked about the most stressful things in their everyday life or at work during the pandemic. Possible infection of family is a major concern that has been reported several times before, for example, in the United Kingdom [38] or formerly among Taiwanese nurses during the SARS outbreak [29] and Chinese health care workers during COVID-19 [30]. Our results confirm the dilemma already mentioned by Perrin et al [19] during SARS: health care workers do their job by helping others but at the same time feel anxious about getting infected or infecting their families. Our participants were less worried about getting infected themselves than infecting their families with COVID-19.

The strategies most frequently used by medical professionals to deal with this unusual situation were "taking protective measures (washing hands, wearing a mask, taking own temperature)" and "actively acquiring more knowledge about COVID-19 (symptoms, transmission pathway, etc)." Effective protective measures were also the most common coping strategy among Taiwanese nurses during SARS [29] and Chinese health care workers during COVID-19 [30]. Another important strategy was "video-chatting with family and friends to share concerns and support," which apparently had a higher priority for the participants in our study when compared to the nurses in Taiwan ("chatting with family and friends by phone to share concerns and support") during SARS [29]. However, nowadays there are more possibilities, especially via social media, to be in touch digitally with friends and family compared to during the SARS outbreak. This has the advantage to get in touch directly with people experiencing mental burden, with the help of so-called e-mental health applications. The increasing role of these web-based interventions during the pandemic has already been observed [46]. While the acceptance of this development, especially among medical professionals, is high [47], different generations follow their own patterns of usage. However, all generations seek to stay related to their family members [48].

COVID-19 and Mental Health in Europe

Our results show that although the majority of respondents reported normal to mild levels of depression, anxiety, and stress, the mean overall level of mental strain experienced was up to 2x higher compared to the normative data means of the DASS-21 [49]. However, according to DASS guidelines, it should be noted here that there is no DASS-21 cut-off for clinical diagnostics [35].

Our results concur with earlier studies about COVID-19 that have reported elevated levels of psychological distress during the pandemic [12-15,50,51]. However, these studies report only about a particular European country, which makes it clear that COVID-19 has a negative effect on the psyche but neglects that there can be differences across countries. The descriptive cross-sectional overview of our study shows that there are differences among countries in the numbers of people belonging to the severe/extremely severe category for depression, anxiety, and stress.

Participants from the United Kingdom and France showed, on a descriptive level, the highest scores for depression, anxiety, and stress when compared to other countries. This may be because England and France were among the countries most affected by COVID-19 [2] with a case fatality rate of 19.2% for France and 14.7% for the United Kingdom by the end of May [5]. Similar to our study, a previous study in the United Kingdom found elevated scores for depression and anxiety during COVID-19 [51], and a study from France presented a considerable prevalence of anxiety 1 week after the start of the lockdown [15]. In Italy and Spain, even though the situation was worse, the participants in our study showed lower scores of psychological strain compared to France and the United Kingdom. One reason for their lower scores of depression, anxiety, and stress could be the high proportion of medical professionals in the Italian sample, whose overall mental health was significantly better than that of the nonmedical professionals. Another reason could be that the surveys started at different points of time in these countries and the peak of the pandemic was different for each country.

The lower levels of psychological distress among participants in Austria and Switzerland could be attributed to the countries' relatively lower number of cases per 1 million people [5]. In Germany, which had a higher number of cases [5] but less psychological strain, the health care system seemed to be better prepared as this is the country with the highest number of critical care beds in Europe [2,52].

Limitations

Although our findings support previous studies on the psychological burden of COVID-19, a few limitations should be considered. Links to the online survey were distributed via social media and via the personal and professional networks of the authors. Since the contact networks in the individual European countries were not equally strong and online distribution of a link was difficult to control, the number of participants for each country was different, leading to uneven distribution of professional groups per country.

Moreover, the surveys did not start simultaneously in all European countries and data could not be acquired when the COVID-19 outbreak peaked in each country. In addition, translating questionnaires into different languages always carries the risk that the individual translations are not completely identical. Since we also partially adapted the already translated versions of DASS-21 to our online format, this could have led to an additional language bias. Finally, the category "nonmedical professionals" was heterogeneous. Persons who worked in nonmedical sectors of the health care system were included in this category and might have been exposed to COVID-19.

Implications

The COVID-19 pandemic has caused fundamental changes in the health care and non–health care sectors and has put considerable strain not only on medical but also on nonmedical professionals. A sizeable part of participants expressed moderate to extremely severe symptoms of depression, anxiety, and/or stress while nonmedical professionals seemed to be more burdened than their medical counterparts. Targeted and personalized mental health services are needed not only for medical professionals but also for other professional groups during pandemics. When developing these services, specific needs and fears should be taken into account. One approach could be to examine the reasons why the medical staff are better at handling the pandemic situation and using these results to develop or optimize mental health services for future pandemics.

By providing the opportunity for medical professionals to carry out their own protective measures and by providing sufficient information about the virus, they might be able to better overcome such situations. Further research is needed to analyze the long-term consequences of the psychological strain of COVID-19 by using valid diagnostic tools and other research designs like longitudinal surveys or qualitative studies. In-depth interviews could provide additional valuable information on major stressors and coping strategies.

Acknowledgments

This study would not have been possible without the support of European health care institutions that helped us distribute the survey link. We thank our Spanish supporters Fundación INTRAS; the Regional Authority of Social Affairs (Castilla y León); the nursing homes ACALERTE and LARES; Hospital San Pedro en Logroño, La Rioja; the Fundación Rey Ardid (Aragón region); and the care organizations Pronisa, AFAVITAE, and ACyLNP. In France, we would like to thank the Marseille Public University Hospital System. In Austria, we thank the Sozialdienste Wolfurt for their support.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19): situation report – 131. World Health Organization. 2020 May 30. URL: https://reliefweb.int/sites/reliefweb.int/files/resources/20200530-covid-19-sitrep-131.pdf [accessed 2020-09-16]
2. Pillai S, Siddika N, Hoque Apu E, Kabir R. COVID-19: Situation of European Countries so Far. Arch Med Res 2020 Oct;51(7):723-725 [FREE Full text] [doi: 10.1016/j.arcmed.2020.05.015] [Medline: 32475614]
3. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ 2020 Aug 10;729:138817 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138817] [Medline: 32360907]
4. Carroll W, Strenger V, Eber E, Porcaro F, Cutrera R, Fitzgerald D, et al. European and United Kingdom COVID-19 pandemic experience: The same but different. Paediatr Respir Rev 2020 Sep;35:50-56 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.012] [Medline: 32709461]
5. Daily new confirmed COVID-19 cases per million people. Our World in Data. 2020 May 30. URL: https://ourworldindata.org/coronavirus [accessed 2020-11-20]
6. Silva C, Cruz C, Torres D, Munuzuri A, Carballosa A, Area I, et al. Optimal control of the COVID-19 pandemic: controlled sanitary deconfinement in Portugal. arXiv Preprint posted online Sept 1, 2020 [FREE Full text]
7. Capacidade de Medicina Intensiva aumentou 23%. Direcao-Geral da Saúde. 2020 Jun 13. URL: https://covid19.min-saude.pt/capacidade-de-medicina-intensiva-aumentou-23/ [accessed 2020-11-20]
8. Decreto del presidente del consiglio dei ministri 08 marzo 2020, Italy. Ministero della Salute. 2020 Mar 8. URL: https://www.trovanorme.salute.gov.it/norme/dettaglioAtto?id=73594 [accessed 2020-09-16]
9. Verordnung 2 über Massnahmen zur Bekämpfung des Coronavirus (COVID-19), Änderung vom 16. Bundesrat Schweiz. 2020 Mar 16. URL: https://www.admin.ch/opc/de/official-compilation/2020/783.pdf [accessed 2020-09-16]
10. Limcaoco R, Mateos E, Fernandez J, Roncero C. Anxiety, worry and perceived stress in the world due to the COVID-19 pandemic, March 2020. Preliminary results. medRxiv Preprint posted online April 6, 2020. [doi: 10.1101/2020.04.03.20043992]
11. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. JAMA Netw Open 2020 Sep 02;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
12. Mazza C, Ricci E, Biondi S, Colasanti M, Ferracuti S, Napoli C, et al. A Nationwide Survey of Psychological Distress among Italian People during the COVID-19 Pandemic: Immediate Psychological Responses and Associated Factors. Int J Environ Res Public Health 2020 May 02;17(9):3165 [FREE Full text] [doi: 10.3390/ijerph17093165] [Medline: 32370116]
13. Odriozola-González P, Planchuelo-Gómez Á, Irurtia MJ, de Luis-García R. Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. Psychiatry Res 2020 Aug;290:113108 [FREE Full text] [doi: 10.1016/j.psychres.2020.113108] [Medline: 32450409]
14. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. The Lancet Psychiatry 2020 Oct;7(10):883-892. [doi: 10.1016/s2215-0366(20)30308-4]
15. Chan-Chee C, Léon C, Lasbeur L. The mental health of the French facing the COVID-19 crisis: Prevalence, Evolution and Determinants of Anxiety Disorders During the First Two Weeks of Lockdown (Coviprev Study, March 23-25 and March 30 – April 1st, 2020). Bull Epidemiol Hebd (Paris). No13ENG 2020 May 7:2-9 [FREE Full text]
16. Cheng SKW, Wong CW, Tsang J, Wong KC. Psychological distress and negative appraisals in survivors of severe acute respiratory syndrome (SARS). Psychol Med 2004 Oct 21;34(7):1187-1195. [doi: 10.1017/s0033291704002272] [Medline: 15697045]

17. Mak IWC, Chu CM, Pan PC, Yiu MGC, Chan VL. Long-term psychiatric morbidities among SARS survivors. Gen Hosp Psychiatry 2009 May;31(4):318-326 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.03.001] [Medline: 19555791]
18. Castelletti S. A shift on the front line. N Engl J Med 2020 Jun 04;382(23):e83. [doi: 10.1056/nejmp2007028]
19. Perrin PC, McCabe OL, Everly GS, Links JM. Preparing for an influenza pandemic: mental health considerations. Prehosp Disaster Med 2009 Jun 28;24(3):223-230. [doi: 10.1017/s1049023x00006853] [Medline: 19618359]
20. Grace SL, Hershenfield K, Robertson E, Stewart DE. The occupational and psychosocial impact of SARS on academic physicians in three affected hospitals. Psychosomatics 2005 Sep;46(5):385-391 [FREE Full text] [doi: 10.1176/appi.psy.46.5.385] [Medline: 16145182]
21. Rossi R, Socci V, Pacitti F, Di Lorenzo G, Di Marco A, Siracusano A, et al. Mental Health Outcomes Among Frontline and Second-Line Health Care Workers During the Coronavirus Disease 2019 (COVID-19) Pandemic in Italy. JAMA Netw Open 2020 May 28;3(5):e2010185. [doi: 10.1001/jamanetworkopen.2020.10185] [Medline: 32463467]
22. Minder R, Peltier E. Virus Knocks Thousands of Health Workers Out of Action in Europe. New York Times. 2020 Mar 24. URL: https://www.nytimes.com/2020/03/24/world/europe/coronavirus-europe-covid-19.html [accessed 2020-09-25]
23. Galbraith N, Boyda D, McFeeters D, Hassan T. The mental health of doctors during the COVID-19 pandemic. BJPsych Bull 2020 Apr 28:1-4 [FREE Full text] [doi: 10.1192/bjb.2020.44] [Medline: 32340645]
24. Maben J, Bridges J. Covid-19: Supporting nurses' psychological and mental health. J Clin Nurs 2020 Jun 02;29(15-16):2742-2750 [FREE Full text] [doi: 10.1111/jocn.15307] [Medline: 32320509]
25. Zhang W, Wang K, Yin L, Zhao W, Xue Q, Peng M, et al. Mental Health and Psychosocial Problems of Medical Health Workers during the COVID-19 Epidemic in China. Psychother Psychosom 2020 Apr 9;89(4):242-250 [FREE Full text] [doi: 10.1159/000507639] [Medline: 32272480]
26. Huang J, Han M, Luo T, Ren A, Zhou X. [Mental health survey of medical staff in a tertiary infectious disease hospital for COVID-19]. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi 2020 Mar 01;38(3):192-195. [doi: 10.3760/cma.j.cn121094-20200219-00063] [Medline: 32131151]
27. Li W, Frank E, Zhao Z, Chen L, Wang Z, Burmeister M, et al. Mental Health of Young Physicians in China During the Novel Coronavirus Disease 2019 Outbreak. JAMA Netw Open 2020 Jun 01;3(6):e2010705 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10705] [Medline: 32478846]
28. Mark G, Smith A. Occupational stress, job characteristics, coping, and the mental health of nurses. Br J Health Psychol 2012 Sep;17(3):505-521. [doi: 10.1111/j.2044-8287.2011.02051.x] [Medline: 22107162]
29. Lee S, Juang Y, Su Y, Lee H, Lin YH, Chao C. Facing SARS: psychological impacts on SARS team nurses and psychiatric services in a Taiwan general hospital. Gen Hosp Psychiatry 2005 Sep;27(5):352-358 [FREE Full text] [doi: 10.1016/j.genhosppsych.2005.04.007] [Medline: 16168796]
30. Cai H, Tu B, Ma J, Chen L, Fu L, Jiang Y, et al. Psychological Impact and Coping Strategies of Frontline Medical Staff in Hunan Between January and March 2020 During the Outbreak of Coronavirus Disease 2019 (COVID-19) in Hubei, China. Med Sci Monit 2020 Apr 15;26:e924171-e924171 [FREE Full text] [doi: 10.12659/MSM.924171] [Medline: 32291383]
31. Du J, Mayer G, Hummel S, Oetjen N, Gronewold N, Zafar A, et al. Mental Health Burden in Different Professions During the Final Stage of the COVID-19 Lockdown in China: Cross-sectional Survey Study. J Med Internet Res 2020 Dec 02;22(12):e24240 [FREE Full text] [doi: 10.2196/24240] [Medline: 33197231]
32. Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare Staff Wellbeing, Burnout, and Patient Safety: A Systematic Review. PLoS One 2016 Jul 8;11(7):e0159015 [FREE Full text] [doi: 10.1371/journal.pone.0159015] [Medline: 27391946]
33. Rothenberger DA. Physician Burnout and Well-Being. Diseases of the Colon & Rectum 2017 Jun;60(6):567-576. [doi: 10.1097/dcr.0000000000000844]
34. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, NSW, Australia: Psychology Foundation of Australia; 1995.
35. Depression Anxiety Stress Scales (DASS). Psychology Foundation of Australia. 2018. URL: http://www2.psy.unsw.edu.au/groups/dass/ [accessed 2020-11-19]
36. Leiner DJ. SoSci Survey (Version 2.6.00-i). URL: https://www.soscisurvey.de [accessed 2020-07-31]
37. IBM SPSS Statistics for Windows, Version 24. Armonk, NY: IBM Corp; 2016.
38. Walton M, Murray E, Christian MD. Mental health care for medical staff and affiliated healthcare workers during the COVID-19 pandemic. Eur Heart J Acute Cardiovasc Care 2020 Apr 28;9(3):241-247 [FREE Full text] [doi: 10.1177/2048872620922795] [Medline: 32342698]
39. Koh D, Lim MK, Chia SE, Ko SM, Qian F, Ng V, et al. Risk perception and impact of Severe Acute Respiratory Syndrome (SARS) on work and personal lives of healthcare workers in Singapore: what can we learn? Med Care 2005 Jul;43(7):676-682. [doi: 10.1097/01.mlr.0000167181.36730.cc] [Medline: 15970782]
40. Maunder R, Hunter J, Vincent L, Bennett J, Peladeau N, Leszcz M, et al. The immediate psychological and occupational impact of the 2003 SARS outbreak in a teaching hospital. CMAJ 2003 May 13;168(10):1245-1251 [FREE Full text] [Medline: 12743065]
41. Mak WW, Law RW, Woo J, Cheung FM, Lee D. Social support and psychological adjustment to SARS: the mediating role of self-care self-efficacy. Psychol Health 2009 Feb;24(2):161-174. [doi: 10.1080/08870440701447649] [Medline: 20186649]

42. Tan BY, Chew NW, Lee GK, Jing M, Goh Y, Yeo LL, et al. Psychological Impact of the COVID-19 Pandemic on Health Care Workers in Singapore. Annals of Internal Medicine 2020 Aug 18;173(4):317-320. [doi: 10.7326/m20-1083]
43. Pieh C, Budimir S, Probst T. Corrigendum to "The effect of age, gender, income, work, and physical activity on mental health during coronavirus disease (COVID-19) lockdown in Austria" [Journal of Psychosomatic Research 136 (2020) 110186]. J Psychosom Res 2020 Dec;139:110278 [FREE Full text] [doi: 10.1016/j.jpsychores.2020.110278] [Medline: 33075602]
44. Rossi R, Socci V, Talevi D, Mensi S, Niolu C, Pacitti F, et al. COVID-19 Pandemic and Lockdown Measures Impact on Mental Health Among the General Population in Italy. Front Psychiatry 2020 Aug 7;11:790 [FREE Full text] [doi: 10.3389/fpsyt.2020.00790] [Medline: 32848952]
45. Jackson D, Anders R, Padula WV, Daly J, Davidson PM. Vulnerability of nurse and physicians with COVID-19: Monitoring and surveillance needed. J Clin Nurs 2020 Jun 10;29(19-20):3584-3587 [FREE Full text] [doi: 10.1111/jocn.15347] [Medline: 32428345]
46. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
47. Mayer G, Gronewold N, Alvarez S, Bruns B, Hilbel T, Schultz JH. Acceptance and Expectations of Medical Experts, Students, and Patients Toward Electronic Mental Health Apps: Cross-Sectional Quantitative and Qualitative Survey Study. JMIR Ment Health 2019 Nov 25;6(11):e14018 [FREE Full text] [doi: 10.2196/14018] [Medline: 31763990]
48. Mayer G, Alvarez S, Gronewold N, Schultz JH. Expressions of Individualization on the Internet and Social Media: Multigenerational Focus Group Study. J Med Internet Res 2020 Nov 04;22(11):e20528 [FREE Full text] [doi: 10.2196/20528] [Medline: 33146622]
49. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
50. Braun M, Niederkrotenthaler T, Till B. SARS CoV-2: Mental Health in Österreich Ausgewählte Ergebnisse zur ersten Befragungswelle. 2020. URL: http://www.suizidforschung.at/wp-content/uploads/2020/05/Mental_Health_Austria_Bericht_Covid19_Welle_I_Mai19_2020.pdf [accessed 2020-09-25]
51. Jia R, Ayling K, Chalder T, Massey A, Broadbent E, Coupland C, et al. Mental health in the UK during the COVID-19 pandemicarly observations. medRxiv Preprint posted online May 19, 2020. [doi: 10.1101/2020.05.14.20102012]
52. Rhodes A, Ferdinande P, Flaatten H, Guidet B, Metnitz PG, Moreno RP. The variability of critical care bed numbers in Europe. Intensive Care Med 2012 Oct 10;38(10):1647-1653. [doi: 10.1007/s00134-012-2627-8] [Medline: 22777516]

Abbreviations

DASS-21: Depression Anxiety Stress Scales–21
MANOVA: multivariate analysis of variance
SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 13.10.20; peer-reviewed by A Bregenzer; comments to author 03.11.20; revised version received 20.11.20; accepted 03.12.20; published 18.01.21.

Please cite as:
Hummel S, Oetjen N, Du J, Posenato E, Resende de Almeida RM, Losada R, Ribeiro O, Frisardi V, Hopper L, Rashid A, Nasser H, König A, Rudofsky G, Weidt S, Zafar A, Gronewold N, Mayer G, Schultz JH
Mental Health Among Medical Professionals During the COVID-19 Pandemic in Eight European Countries: Cross-sectional Survey Study
J Med Internet Res 2021;23(1):e24983
URL: http://www.jmir.org/2021/1/e24983/
doi:10.2196/24983
PMID:33411670

Original Paper

Assessment of Internet Hospitals in China During the COVID-19 Pandemic: National Cross-Sectional Data Analysis Study

Xingyan Xu[1*], MMed; Yingying Cai[1*], MMed; Siying Wu[1*], MD; Jianhui Guo[1], MMed; Le Yang[1], MMed; Jieli Lan[1], MPH; Yi Sun[2], MD; Bingbing Wang[1], BMed; Jieyu Wu[1], MPH; Tinggui Wang[1], MPH; Shuna Huang[3], MMed; Yawen Lin[1], MMed; Yuduan Hu[1], MMed; Mingjun Chen[1], MMed; Xuecai Gao[1], BMed; Xiaoxu Xie[1], MD

[1]Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, Fuzhou, China
[2]Department of Preventive Medicine, School of Public Health, Fujian Medical University, Fuzhou, China
[3]Department of Clinical Research and Translation Center Office, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China
[*]these authors contributed equally

Corresponding Author:
Xiaoxu Xie, MD
Department of Epidemiology and Health Statistics
School of Public Health
Fujian Medical University
No. 1 Xuefu North Road
Minhou County
Fuzhou, 350122
China
Phone: 86 059122862023
Email: xiexiaoxu@aliyun.com

Abstract

Background: Internet hospitals in China are being rapidly developed as an innovative approach to providing health services. The ongoing COVID-19 pandemic has triggered the development of internet hospitals that promote outpatient service delivery to the public via internet technologies. To date, no studies have assessed China's internet hospitals during the COVID-19 pandemic.

Objective: This study aimed to elucidate the characteristics of China's internet hospitals and assess the health service capacity of these hospitals.

Methods: Data on 711 internet hospitals were collected from official websites, the WeChat (Tencent Inc) platform, smartphone apps, and the Baidu search engine until July 16, 2020.

Results: As of July 16, 2020, 711 internet hospitals were developed in mainland China. More than half of these internet hospitals (421/711, 59.2%) were established during 2019 (206/711, 29%) and 2020 (215/711, 30.2%). Furthermore, about one-third (215/711, 30.2%) of internet hospitals were established at the beginning of 2020 as an emergency response to the COVID-19 epidemic. The 711 internet hospitals consisted of the following 3 types of hospitals: government-oriented (42/711, 5.91%), hospital-oriented (143/711, 20.11%), and enterprise-oriented internet hospitals (526/711, 73.98%). The vast majority of internet hospitals were traditional hospitals (526/711, 74%). Nearly 46.1% (221/711) of internet hospitals requested doctors to provide health services at a specific web clinic. Most patients (224/639, 35.1%) accessed outpatient services via WeChat. Internet hospitals' consulting methods included SMS text messaging consultations involving the use of graphics (552/570, 96.8%), video consultations (248/570, 43.5%), and telephone consultations (238/570, 41.8%). The median number of available web-based doctors was 43, and the median consultation fees of fever clinics and other outpatient clinics were ¥0 (US $0) per consultation and ¥6 (US $0.93) per consultation, respectively. Internet hospitals have provided various services during the COVID-19 pandemic, including medical prescription, drug delivery, and medical insurance services.

Conclusions: The dramatic increase of internet hospitals in China has played an important role in the prevention and control of COVID-19. Internet hospitals provide different and convenient medical services for people in need.

(J Med Internet Res 2021;23(1):e21825) doi:10.2196/21825

KEYWORDS

internet hospital; COVID-19; prevention; control; health care; China; cross-sectional; digital health; accessibility

Introduction

The outbreak of the novel COVID-19 disease was first reported in January 2020, and the disease has been spreading throughout the entire globe without any indication of stopping any time soon [1,2]. This epidemic has constituted a public health emergency of pandemic proportions [3], which has created major dilemmas in all areas of health care. Such dilemmas include a lack of hospital resources, the suspension of outpatient services, inconvenience in seeking medical treatment due to social isolation, and lockdown. In addition, the human-to-human transmission of the SARS-CoV-2 virus has become the primary transmission route of COVID-19 [4].

Strict prevention measures and effective therapeutics are urgently needed for the effective control of the COVID-19 pandemic. Such measures include the promulgation of Chinese government decrees that require domestic internet hospitals to vigorously develop telemedicine services [5]. Currently, internet hospitals have been developed as a new approach for the provision of health services, including outpatient services delivered through the internet or various digital technologies [6,7]. Internet hospitals surmount geographical obstacles and time-related barriers, thereby making it easy for Chinese patients to rapidly seek doctor services [8]. Additionally, internet hospitals reduce the chance of nosocomial cross infection by potentially keeping patients away from hospitals. This protects both clinicians and other patients, which is a practice that is welcomed by most Chinese people [9]. Internet hospitals have proliferated as Chinese people continue to seek health services and psychological assistance amid the COVID-19 pandemic.

China's new internet hospitals, which were established during the epidemic, have yet to be systematically and exhaustively assessed. In this study, we aimed to provide an overview of internet hospitals in China (ie, as of July 2020) through a national cross-sectional study that describes the characteristics of these hospitals. We also aimed to evaluate the health service capacity of internet hospitals during the COVID-19 epidemic.

Methods

Internet Hospital Eligibility

Data were collected by conducting a search on national websites, including the popular Baidu search engine and the official websites of the National Health Commission of the People's Republic of China and each provincial health commission. The search was conducted by using specific Chinese search terms, such as those for "Internet hospitals," "Internet health," "Internet medicine," "mobile medicine," "mobile health," "Telehealth," "digital medicine," "digital health," "Web hospitals," and "Cloud hospitals." All officially registered internet hospitals were eligible for inclusion. Data from October 10, 2006 to July 15, 2020 were collected. Additionally, we used a public WeChat (Tencent Inc) account and app stores (ie, the Apple App Store, Android App Store, Huawei App Store, Xiaomi App Store, and 360 App Store) to find medical care platforms and mobile health apps that were affiliated with internet hospitals. The study region included all of mainland China, except Hong Kong, Macau, Taiwan, and several islands in the South China Sea. Ultimately, we found 711 internet hospitals.

Data Collection

We searched for information about each internet hospital by using official websites, the WeChat platform (ie, one of the largest mobile social network apps in China, with more than 1 billion monthly active users), and smartphone apps (ie, if available). Collected data were recorded on a digital archive via Microsoft Excel. Information on internet hospital features included construction date, location, the identity of investors, hospital characteristics, and the role that the hospital played in combating COVID-19.

We assessed the consultation characteristics of internet hospitals, including access method (ie, website, app, WeChat, and other); consulting method (ie, video consultations, telephone consultations, and SMS text messaging consultations involving the use of both text and graphics); medical insurance, prescription, and drug delivery service provision; doctor source (ie, local hospital, medical union, district, and nationwide doctors); service time (ie, any time or specific times); the number of available web-based doctors who could deliver instant medical services; and the availability of outpatient departments.

In this study, the roles played by internet hospitals referred to whether the hospital provided COVID-19–related inquiries regarding fever clinic consultations, psychological counseling, the provision of COVID-19–related information, and whether there was an overseas version of an internet hospital that could help in the fight against COVID-19. When collecting information on fever clinics, we focused on how consultations were conducted (ie, manual, artificial intelligence, or both manual and artificial intelligence) and the fee per consultation. Additionally, we collected information on internet hospital–related policies and regulations that were released by the government, management agencies, medical institutions, and health care industry associations. This helped us to understand the actions taken by the health community in China during the COVID-19 outbreak and assess the role of internet hospitals in combating COVID-19. At least 2 researchers extracted data for each internet hospital independently. Any objections to the data were resolved via discussion.

Data Analysis and Visualization

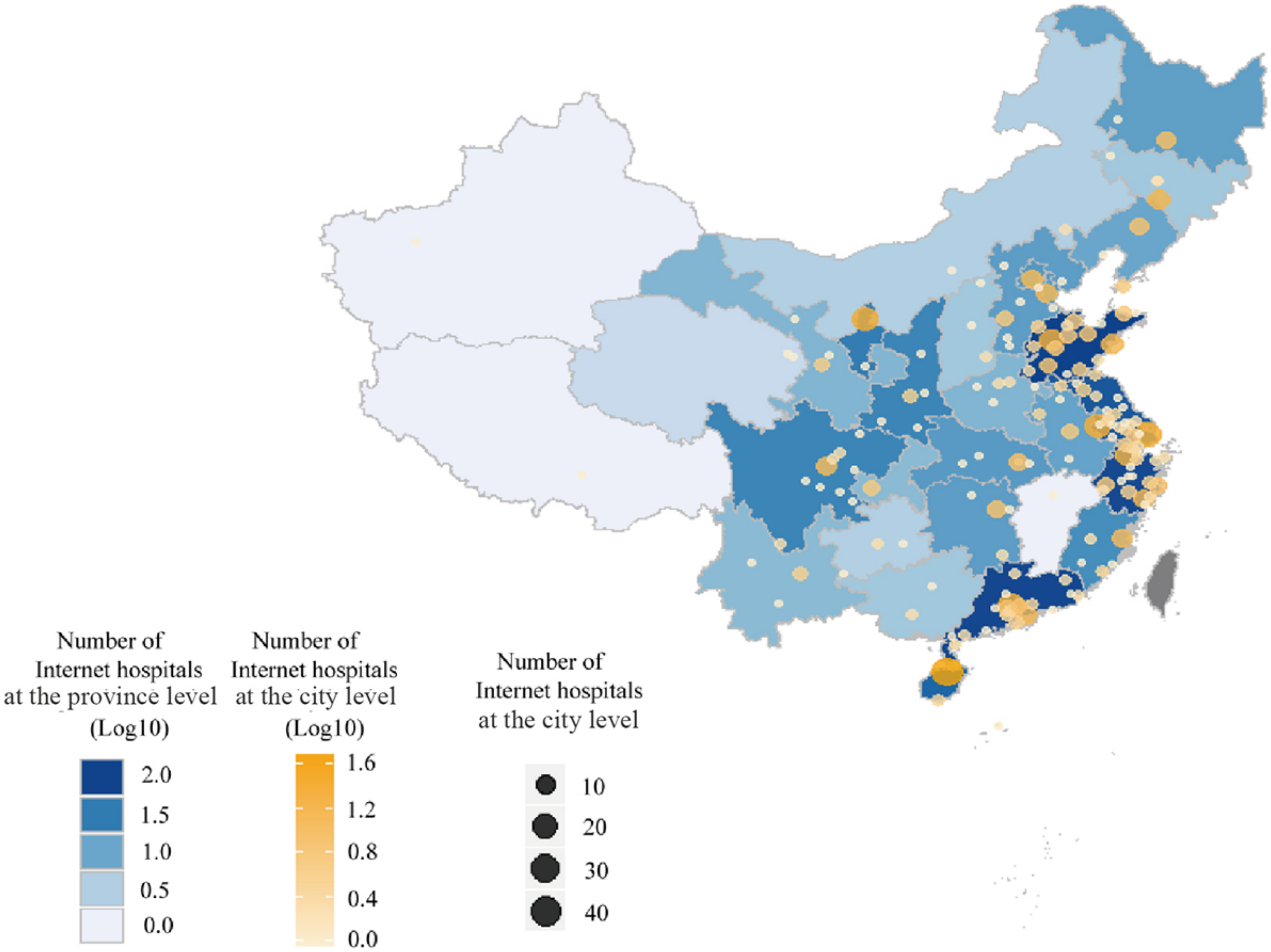
Data management and analysis were performed with Microsoft Excel 2017. Geographical and time distributions were drawn with R 4.0.2. Count data were expressed as number (percentage) and presented on bar charts, Venn diagrams, mosaic plots, Nightingale rose diagrams, and statistical maps. Skewed continuous data were expressed as median (interquartile range) and presented on Violin plots.

Results

Internet Hospital Characteristics

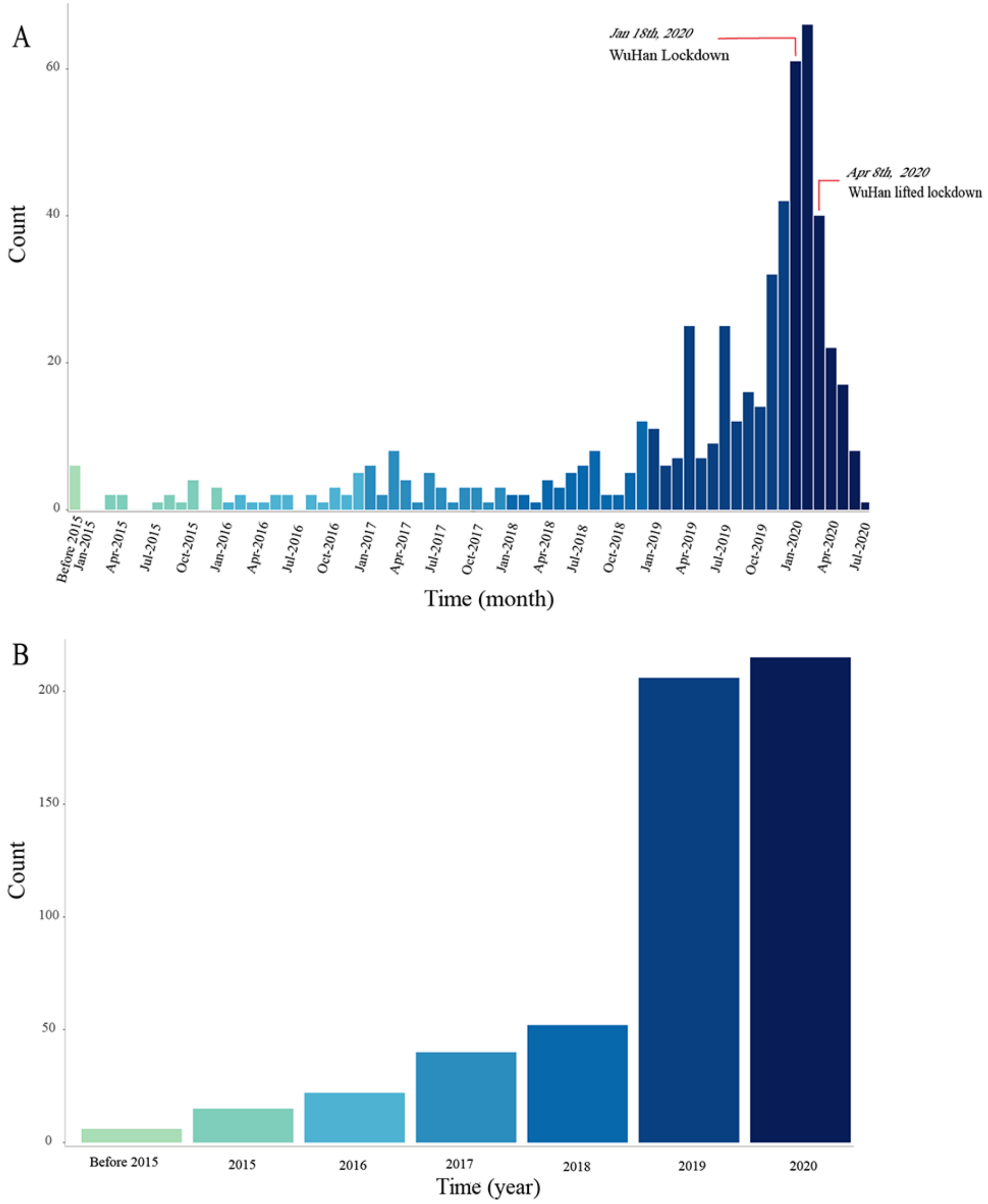
As of July 2020, data on 711 internet hospitals were collected. Generally, internet hospitals were mainly distributed in the east and southern coastal provinces (Figure 1). This distribution was closely related to the foundation and early development of internet hospitals in these areas [10]. However, internet hospitals were not limited to these first-to-try areas. At the time of this study, every province had at least 1 internet hospital. This reflects the trend of gradual development from coastal cities to inland regions (ie, the gradual spread of internet hospitals from the point of origin to mainland China).

Figure 1. Spatial distributions of internet hospitals in China. The different shades of blue denote the total number of internet hospitals in each province. The color (ie, the different shades of yellow) and size of the circles indicate the number of internet hospitals in each city.



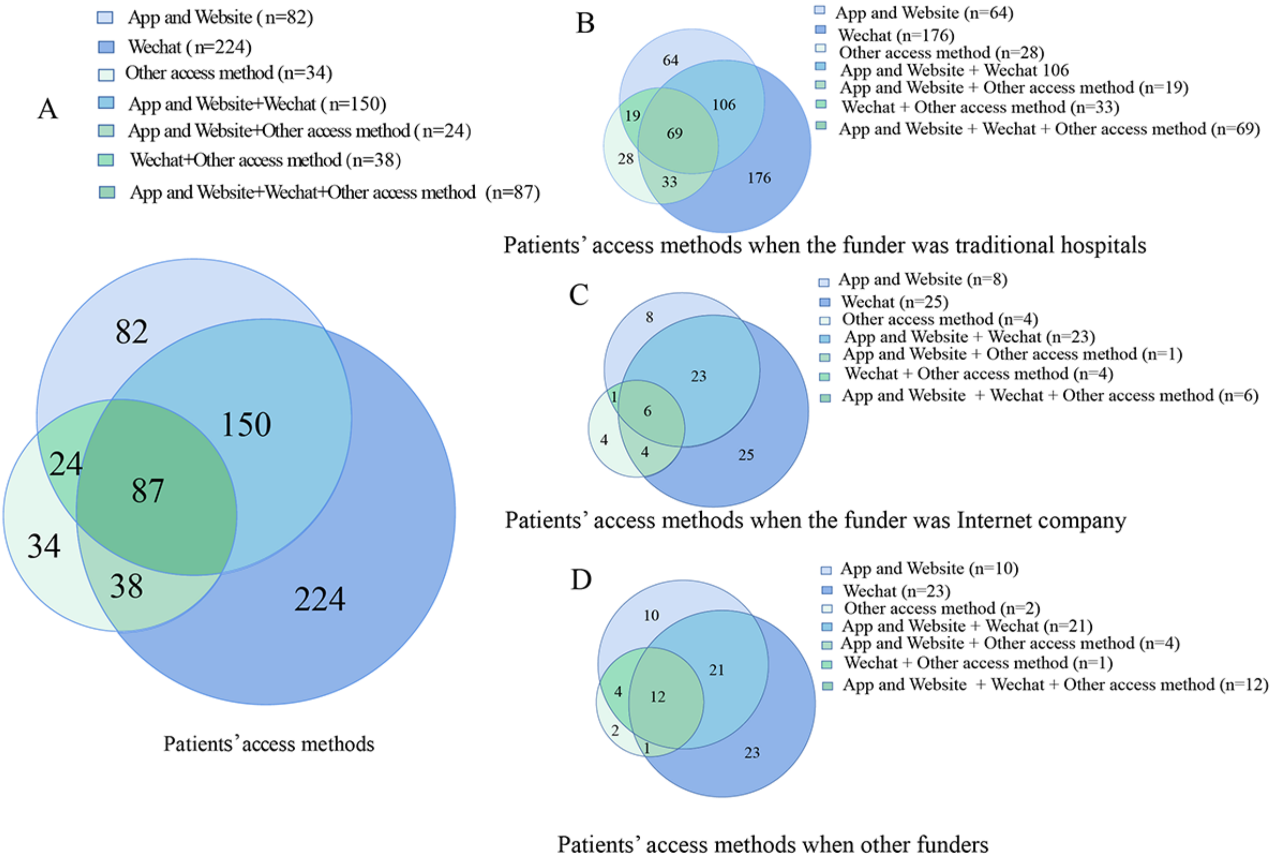
In recent years, internet hospitals have been rapidly developed. Of the 711 internet hospitals found in this study, 215 (30.1%) were established in 2020 (ie, after the beginning of the COVID-19 outbreak) and 5 (0.7%) were established before 2015 and until July 2020. Moreover, 2.1% (15/711), 3.1% (22/711), 5.6% (40/711), 7.3% (52/711), 29% (206/711) and 30.2% (215/711) of internet hospitals were established in 2015, 2016, 2017, 2018, 2019, and 2020, respectively (Figure 2; Table S2 in Multimedia Appendix 1). Notably, we could not identify the date of foundation for 21.9% (156/711) of internet hospitals. We concluded that there has been a growing trend for the establishment of internet hospitals up to July 2020. Additionally, the government has issued a series of policies that were conducive to the role of internet hospitals in combating COVID-19 (Figure 2).

Figure 2. The construction dates of internet hospitals in China. (A) The construction dates of internet hospitals in China by month. (B) The construction dates of internet hospital in China by year. The following important policies were issued during the COVID-19 epidemic in China: (1) Notice on internet diagnosis and treatment consultation services for COVID-19 prevention and control (February 6, 2020), (2) Notice on strengthening internet diagnosis and treatment consulting services for COVID-19 prevention and control (February 8, 2020), (3) Notice on the national telemedicine and internet medical center for the national remote consultation of critically ill patients with COVID-19 (February 21, 2020), (4) Notice on further promoting the development and standardized management of internet medical services (April 18, 2020), and (5) Notice on the technical specifications and financial management of the "Internet+ medical service" project of public medical institutions (May 8, 2020).



In terms of patients' (N=639) methods for accessing internet hospitals, 13.6% (87/639) of patients accessed internet hospitals by using apps, websites, WeChat, and other access methods simultaneously. Of note, 12.8% (82/639) of patients visited internet hospitals by exclusively using apps and websites, 35.1% (224/639) visited internet hospitals by exclusively using WeChat, and 5.3% (34/639) visited internet hospitals by using other access methods. Additionally, regardless of who funded the internet hospitals, the most widely used service that patients used to access outpatient services was WeChat, followed by apps or websites, and other access methods (Figure 3).

Figure 3. Methods for accessing internet hospitals. (A) Patients' access methods. (B) Patients' access methods when the funder was a traditional hospital. (C) Patients' access methods when the funder was an internet company. (D) Patients' access methods when the funder was another type of company.



Patients' access methods

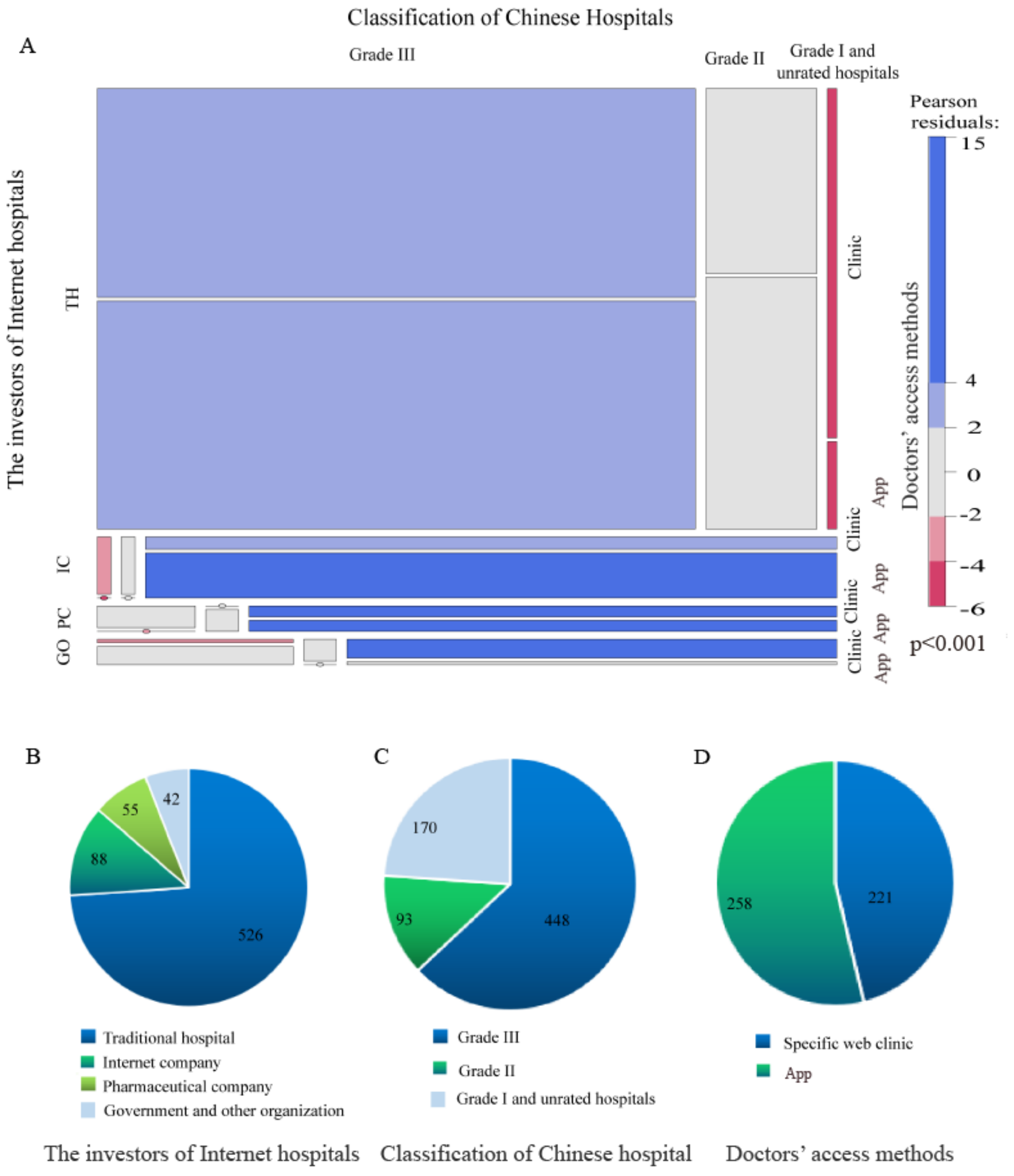
Patients' access methods when the funder was traditional hospitals

Patients' access methods when the funder was Internet company

Patients' access methods when other funders

The vast majority of internet hospitals were traditional hospitals (526/711, 74%). This finding shows that traditional hospitals were more capable of integrating internet hospital services into their infrastructure than other types of hospitals. Grade III hospitals, which have a large number of outpatients, superb medical technology, diverse and complex service types, and various business process requirements, accounted for a large proportion (448/711, 62%) of internet hospitals. Furthermore, government departments, such as health administrations, also supported the construction of internet hospitals. In terms of the consultation characteristics of the 711 internet hospitals, 221 (46.14%) hospitals asked doctors to provide medical services at a specific web clinic, whereas 258 (53.86%) adopted the use of apps for health service delivery (Figure 4).

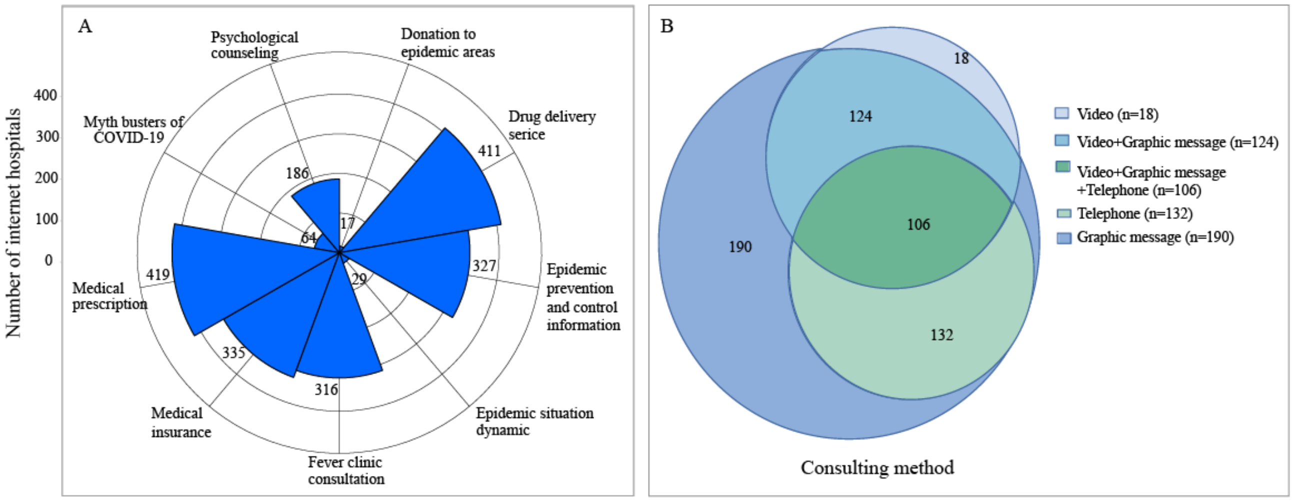
Figure 4. Characteristics of internet hospitals. (A) A mosaic plot with areas that show the proportion of doctors' web-based methods for accessing internet hospitals in China, stratified by the classification of Chinese hospitals and the structure of the main investment. The investors of internet hospitals include THs, ICs, PCs, and GOs. The classification of Chinese hospitals were as follows: Grade III, Grade II, and Grade I and unrated hospitals. Doctors provided medical services through a specific web clinic or app. (B) The investors of internet hospitals. (C) Classification of Chinese hospitals. (D) Doctors' access methods. GO: government and other organization; IC: internet company; PC: pharmaceutical company; TH: traditional hospital.



Medical Service Characteristics

With regard to the roles that the 711 internet hospitals played in combating COVID-19, most hospitals (419/484, 86.6%) provided medical prescriptions. Furthermore, 74.5% of hospitals (411/552) provided drug delivery services, 67.5% (335/496) provided medical insurance, 60.7% (335/496) provided epidemic prevention and control information, 60% (316/527) provided fever clinic consultations, 35.2% (186/529) provided psychological counseling, 13.3% (64/483) provided a myth busters of COVID-19 service, 20.6% (29/141) assessed epidemic situation dynamics, and 3.8% (17/443) made donations (Figure 4). Internet hospitals' consulting methods included SMS text messaging consultations involving the use of graphics (552/570, 96.8%), video consultations (248/570, 43.5%), and telephone consultations (238/570, 41.8%). Moreover, 18.6% (106/570) of internet hospitals offered 3 methods for patient consultations (Figure 5). It should be noted that several internet hospitals were missing pertinent data. As such, these hospitals were excluded from analysis.

Figure 5. Medical service characteristics of China's internet hospitals. (A) The roles that internet hospitals played in fighting COVID-19. (B) Internet hospitals' consulting methods.



Human Medical Resources

The distribution of the internet hospitals' human medical resources in China is shown in Figure 6. The median number of available web-based doctors was 43 (IQR 3-172; maximum: 563,881). The median consultation fees of fever clinics and other outpatient clinics (ie, excluding fever clinics) were ¥0 (US $0; IQR ¥0 [US$ 0]-¥0 [US $0]) per consultation and ¥6 (US $0.93; IQR ¥0 [US $0]-¥20 [US $3.09]) per consultation, respectively (Figure 6, Table 1). About half of the hospitals (204/422, 48.3%) provided services at any time. Furthermore, 80.9% (467/577) of doctors were from local hospitals and medical unions (Figure 6).

Figure 6. Human medical resources in China's internet hospitals. (A) Violin plots that show the number of doctors in China's internet hospitals. The numbers on the y-axis are logarithms to the base 2. (B, C) Violin plots displaying the consultation fees for a person who visited an internet hospital to find medical help (ie, consultation fees for fever clinics and other clinics). (D, E) Bar plots that show the distribution of doctors' service times and doctor sources with regard to internet hospitals, respectively.

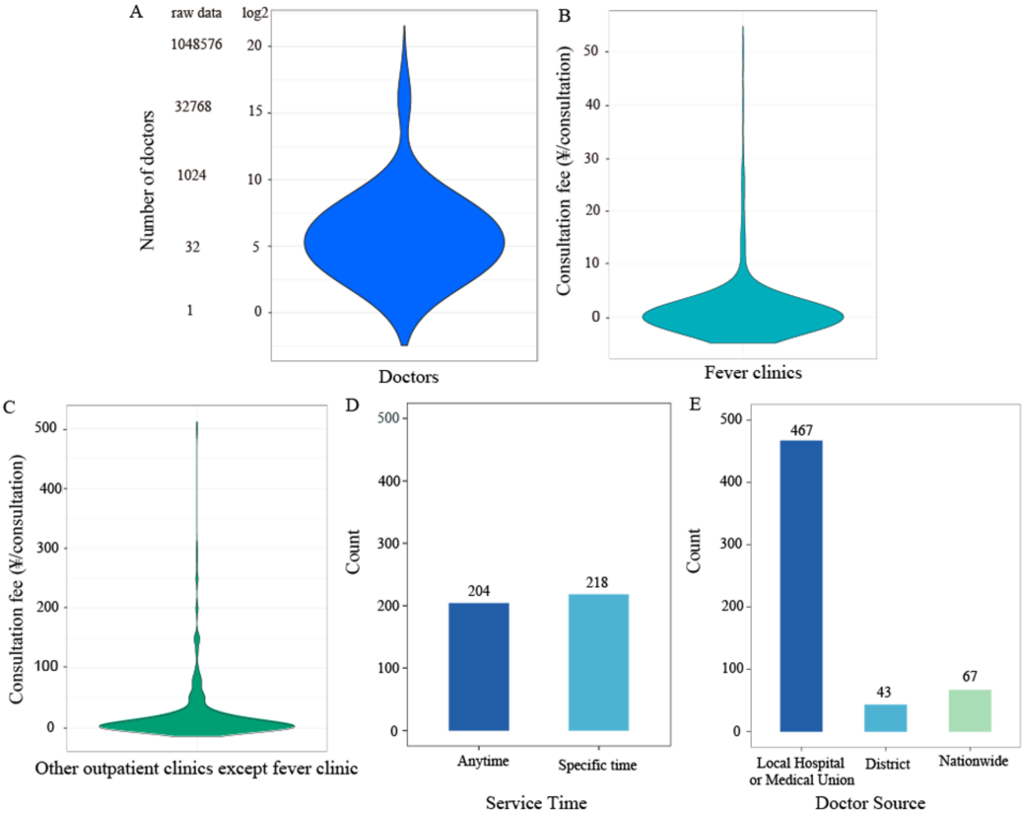


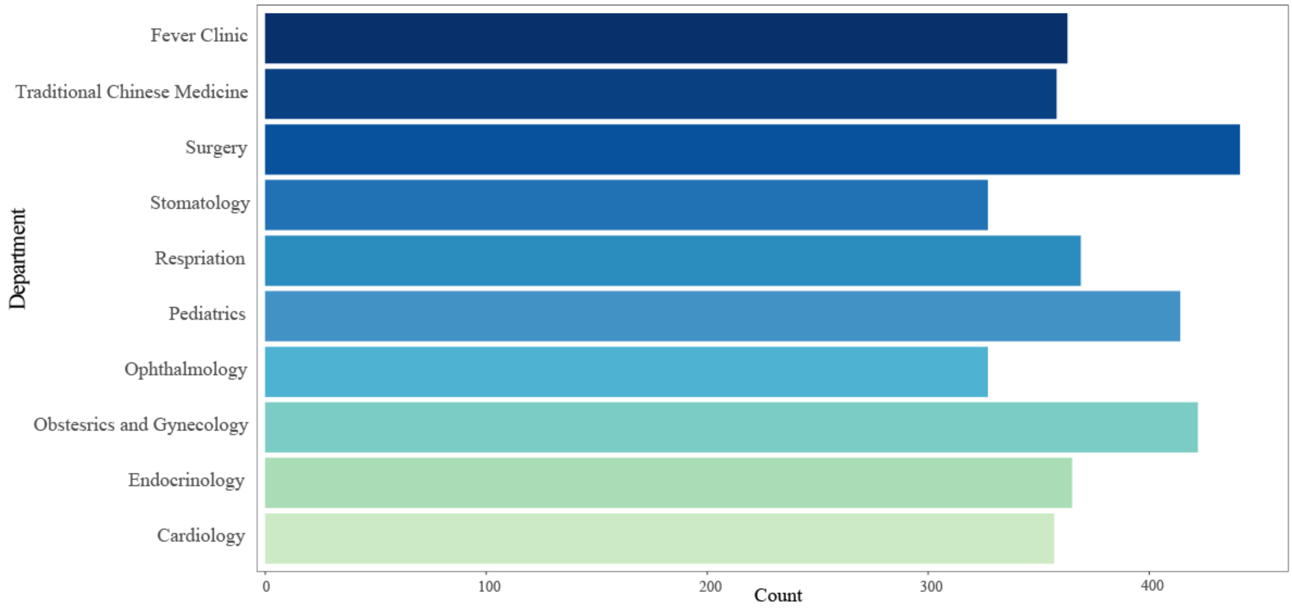
Table 1. Number of doctors and consultation fees for internet hospitals in China.

Variable	Minimum	5th percentile	25th percentile	50th percentile	75th percentile	95th percentile	Maximum
Number of available web-based doctors, n	1	3	11	43	172	1496.4	563881
Consultation fee of fever clinics, ¥[a]/consultation	0	0	0	0	0	8.9	50
Consultation fee of other outpatient clinics (ie, excluding fever clinics), ¥/consultation	0	0	0	6	20	120	500

[a] A currency exchange rate of ¥1=US $0.15 is applicable.

Multiple internet hospitals had different departments for patients; the department of surgery (441/711, 62%) was the most common. In contrast, the number of stomatology (327/711, 46%) and ophthalmology (327/711, 46%) departments was relatively small. Of the 711 internet hospitals, 363 (51.1%) introduced fever clinics in response to the COVID-19 pandemic (Figure 7).

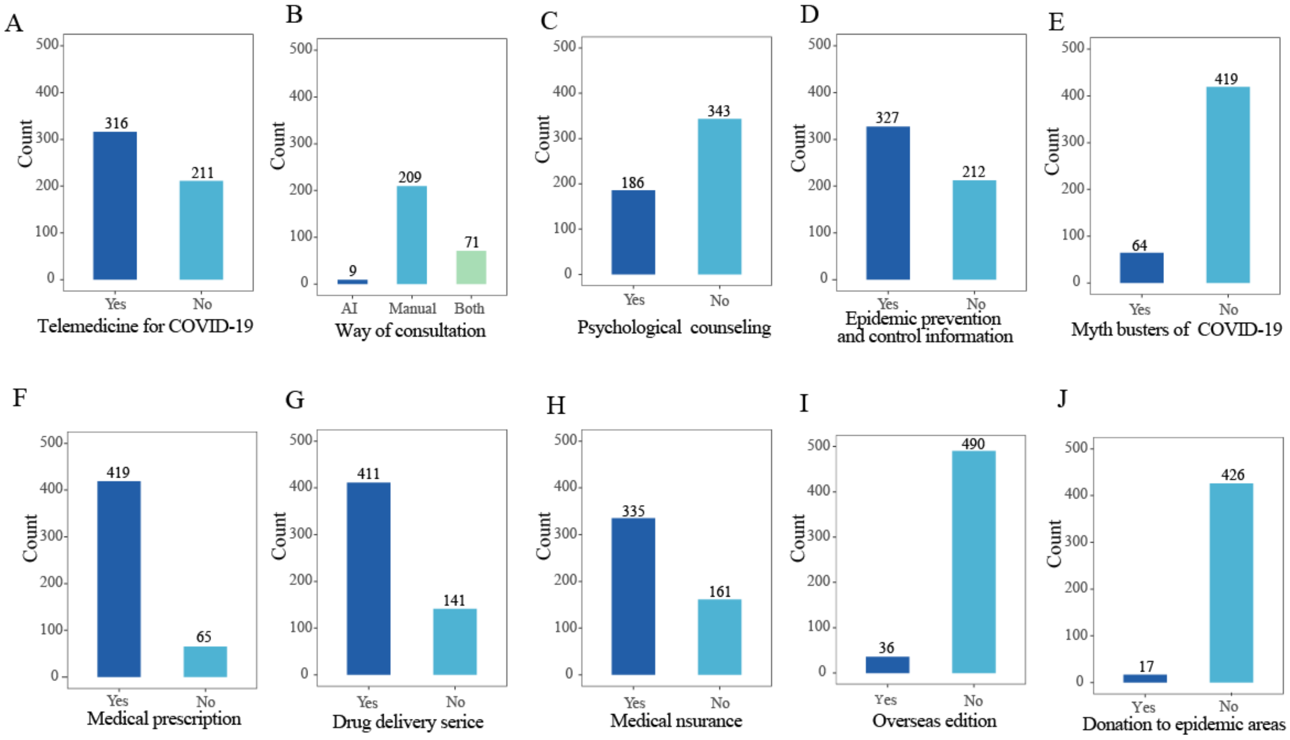
Figure 7. A bar plot that depicts the number of available departments in internet hospitals.



Relation to COVID-19

With regard to the internet hospitals with available data, 60% (316/527) of internet hospitals provided telemedicine services for managing COVID-19. Internet hospitals mainly conducted manual consultations (209/289, 72.3%). In addition, 35.2% (186/529), 60.7% (327/539), 13.3% (64/483), and 3.8% (17/443) of internet hospitals provided psychological counseling, epidemic prevention and control information, myth busters services, and donations, respectively, in response to COVID-19. Furthermore, a number of internet hospitals supported medical prescription (419/484, 86.6%), drug delivery (411/552, 74.5%), and medical insurance (335/496, 67.5%) services to further enhance COVID-19 prevention and control (Figure 8).

Figure 8. (A-J) Bar plots that depict the roles of internet hospitals during the COVID-19 epidemic. AI: artificial intelligence.



Discussion

Principal Findings

This national cross-sectional study outlines the status of the internet hospitals established in China by July 2020. There has been a rising trend in the emergence of internet hospitals in China. This trend peaked in December 2019 after the China National Healthcare Security Administration promulgated the Guidance of the State Medical Security Bureau on improving the Internet+ medical service price and medical insurance payment policy [11]. Furthermore, the number of internet hospitals peaked in February 2020. This may have been the period with the largest number of internet hospitals established in a single month due to the epidemic.

Due to inequalities in medical resources, medical technology, and informatization level across the country, the status of internet hospitals in China varies greatly. Overall, the eastern and southern coastal provinces have the highest distribution of major internet hospitals. This distribution is closely related to the foundation and early development of internet hospitals in these regions [10]. However, internet hospitals are no longer limited to these areas. Jiangxi, Jilin, and other areas established the first batch of internet hospitals in 2020. This reflects the trend of gradual development from coastal cities to inland areas.

The vast majority of internet hospitals have been constructed by Grade III hospitals, which have a large number of patients, superb medical technology, and diverse and complex health services. The independent construction of an internet hospital system can meet the needs of individualized development. Moreover, the health administration department has been promoting the construction of internet hospitals. A concern however is that in most provinces, traditional hospitals make up a large proportion of internet hospitals; only Ningxia and Hainan are dominated by corporate-led internet hospitals.

The catalytic effect of the COVID-19 pandemic on the internet hospital industry has been obvious since 2020; the pandemic has resulted in major development opportunities for internet hospitals. Due to the need for epidemic prevention and control, the National Health Commission, the National Medical Insurance Bureau, and other health departments have intensively introduced a series of policies [12-16] to vigorously promote the development of internet hospitals. During the epidemic, internet hospitals have launched web-based fever clinics; chronic disease follow-up consultations, prescriptions, and drug delivery services; and several medical insurance reimbursement programs, and these have been continuously and urgently approved. Web-based medical treatment has promoted the public awareness of internet hospitals. Additionally, the epidemic has prompted people to develop a habit of using internet hospitals, to a certain extent [17].

Internet hospitals can reduce crowd gathering in physical hospitals through a variety of methods. Web-based medical services, such as web-based education, publicity, and psychological intervention, can not only help the public master basic anti-epidemic knowledge and skills, but also reduce social panic, thereby reducing the number of unnecessary physical hospital visits and enhancing psychological resilience [18]. By integrating web-based resources with offline epidemic control measures, internet hospitals can play a greater role in epidemic prevention and control [19].

With the launch of internet hospitals, the tension regarding offline medical resources has, to a certain extent, been relieved. With regard to upper first-class hospitals, the most direct role of internet hospitals has been shifting to providing returning patients with web-based services and improving the efficiency of medical treatment. Small and medium regional hospitals with a small number of outpatients or limited construction and operation capabilities have the choice of relying on regional internet hospitals or third-party, internet-based medical platforms. This approach is meant to reduce input costs by making full use of the advantages of internet hospitals, as internet-based platforms can help doctors concentrate on becoming the core resource for serving patients. As a result, internet hospitals and internet-based platforms will achieve the advantage of complementation and achieve a win-win situation.

After several years of development, China's internet hospitals can now overcome the limitations of time and space and provide various medical services with a high degree of accessibility for all patients [20]. Moreover, interdisciplinary and transregional exchanges and cooperation can be achieved with internet hospitals, thereby improving doctors' ability to deal with emerging diseases. The Chinese government encouraged internet hospitals to join epidemic prevention and control efforts at the beginning of the COVID-19 outbreak [11], and confirmed that internet hospitals are an important part of the joint epidemic prevention and control system [21]. To make better use of internet hospitals, more measures should be implemented, such as encouraging more doctors (ie, especially psychologists and pediatricians) to take part in web-based services, discovering the needs of the public in time to adjust response strategies, and developing more standardized consulting service guidelines. At the same time, there is a need to improve the availability of internet hospital services in a variety of ways, strengthen publicity efforts for expanding the internet hospital user base, and cooperate with community service centers and the Centers for Disease Control and Prevention to improve joint epidemic control and prevention mechanisms.

Conclusion

This study demonstrates how the concepts of medical service provision have been redefined by internet hospitals and internet-based technologies. Internet hospitals allow for the transferring of patients to medical centers that can provide medical services and health care without requiring people to leave their homes. Internet hospitals can provide different types of medical services to people seeking medical needs; offer essential medical support to the public during the COVID-19 outbreak, thereby regulating social psychology and easing social panic; maintain proper social distancing; and promote correct medical-seeking behaviors. These advantages result in the reduction of the incidence of cross-infection in hospitals. Thus, it is evident that internet hospitals play an important role in COVID-19 prevention and control.

Acknowledgments

The authors thank all the participants who made this study possible. This study was funded by Fujian medical university talent research funding (Grant No. XRCZX2019031), which was received by XXie.

Authors' Contributions

XXie and SW contributed to the study design. XXu and YC performed statistical analyses, interpreted the data, and drafted the manuscript. XXu, YC, JG, JL, BW, JW, TW, YL, YH, MC, and XG contributed to data collection. LY, YS, and SH revised the manuscript. All authors contributed to the critical revision of the final manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 20 KB - jmir_v23i1e21825_app1.docx]

References

1. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
2. Xu X, Wu X, Jiang X, Xu K, Ying L, Ma C, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ 2020 Feb 19;368:m606 [FREE Full text] [doi: 10.1136/bmj.m606] [Medline: 32075786]
3. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
4. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
5. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]

6. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015 Aug;3(8):e445-e446 [FREE Full text] [doi: 10.1016/S2214-109X(15)00042-X] [Medline: 26187488]
7. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet Hospitals in China: Cross-Sectional Survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
8. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
9. Wu J, Xie X, Yang L, Xu X, Cai Y, Wang T, et al. Mobile health technology combats COVID-19 in China. J Infect. Epub ahead of print 2020 Jul 27 [FREE Full text] [doi: 10.1016/j.jinf.2020.07.024] [Medline: 32730998]
10. He C, Zhou Q, Chen W, Tian J, Zhou L, Peng H, et al. Using an Internet-Based Hospital to Address Maldistribution of Health Care Resources in Rural Areas of Guangdong Province, China: Retrospective and Descriptive Study. JMIR Med Inform 2018 Dec 21;6(4):e51 [FREE Full text] [doi: 10.2196/medinform.9495] [Medline: 30578195]
11. Guidance on improving Internet+ medical service price and medical insurance payment policy. National Healthcare Security Administration. 2019 Aug 30. URL: http://www.nhsa.gov.cn/art/2019/8/30/art_37_1707.html [accessed 2021-01-11]
12. Notice on accomplishing online consultation services in the epidemic prevention and control. National Health Commission of the People's Republic of China. 2020 Feb 02. URL: http://www.nhc.gov.cn/yzygj/s7653p/202002/ec5e345814e744398c2adef17b657fb8.shtml [accessed 2021-01-11]
13. Strengthening the consultation service of Internet diagnosis and treatment in epidemic prevention and control. National Health Commission of the People's Republic of China. 2020 Feb 08. URL: http://www.nhc.gov.cn/yzygj/s7652m/202002/32c3e98988894fa18280e4543d2710c7.shtml [accessed 2021-01-11]
14. Notice on further promoting the development and standardized management of internet medical services. National Health Commission of the People's Republic of China. 2020 May 08. URL: http://www.nhc.gov.cn/yzygj/s3594q/202005/fddc84a0c623460c89091d92c57385fe.shtml [accessed 2021-01-11]
15. Notice on the technical specification and financial management of the Internet plus medical service project in public medical institutions. National Health Commission of the People's Republic of China. 2020 May 13. URL: http://www.nhc.gov.cn/caiwusi/s7788c/202005/43367d28fc7544db855e218f9b26f137.shtml [accessed 2021-01-11]
16. Notice on novel coronavirus pneumonia critical national patients' remote consultation at national telemedicine and Internet Medical Center. National Health Commission of the People's Republic of China. 2020 Feb 21. URL: http://www.nhc.gov.cn/yzygj/s7658/202002/69b24672365043eebc379c8bab30c90d.shtml [accessed 2021-01-11]
17. Berger ZD, Evans NG, Phelan AL, Silverman RD. Covid-19: control measures must be equitable and inclusive. BMJ 2020 Mar 20;368:m1141. [doi: 10.1136/bmj.m1141] [Medline: 32198146]
18. Ho CS, Chee CY, Ho RC. Mental Health Strategies to Combat the Psychological Impact of COVID-19 Beyond Paranoia and Panic. Ann Acad Med Singap 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]
19. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using Mobile Apps for Health Management: A New Health Care Mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
20. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: An examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
21. Notice on launching online services to further strengthen epidemic prevention and control in Hubei. The State Council's Comprehensive Team for Joint Prevention and Control of New Coronavirus Pneumonia. 2020 Feb 26. URL: http://www.gov.cn/xinwen/2020-02/27/content_5483977.htm [accessed 2021-01-11]

Edited by G Eysenbach; submitted 29.09.20; peer-reviewed by L Ding, W Zhang; comments to author 24.10.20; revised version received 14.11.20; accepted 03.12.20; published 20.01.21.

Please cite as:
Xu X, Cai Y, Wu S, Guo J, Yang L, Lan J, Sun Y, Wang B, Wu J, Wang T, Huang S, Lin Y, Hu Y, Chen M, Gao X, Xie X
Assessment of Internet Hospitals in China During the COVID-19 Pandemic: National Cross-Sectional Data Analysis Study
J Med Internet Res 2021;23(1):e21825
URL: http://www.jmir.org/2021/1/e21825/
doi:10.2196/21825
PMID:33417586

Original Paper

Use of a Remote Oncology Pharmacy Service Platform for Patients With Cancer During the COVID-19 Pandemic: Implementation and User Acceptance Evaluation

Zhuo-Jia Chen[1], PhD; Wei-Ting Liang[1], MS; Qing Liu[1], BSc; Rong He[2], ME; Qian-Chao Chen[1], BSc; Qiu-Feng Li[1], BSc; Yao Zhang[1], BSc; Xiao-Dong Du[1], BSc; Ying Pan[1], PhD; Shu Liu[1], PhD; Xiao-Yan Li[1], MS; Xue Wei[1], MS; He Huang[3], MD; Hong-Bing Huang[1*], MS; Tao Liu[1*], MS

[1]Department of Pharmacy, Sun Yat-sen University Cancer Center, Guangzhou, China
[2]Department of Information, Sun Yat-sen University Cancer Center, Guangzhou, China
[3]Department of Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China
[*]these authors contributed equally

Corresponding Author:
Hong-Bing Huang, MS
Department of Pharmacy
Sun Yat-sen University Cancer Center
Guangzhou
China
Email: huanghb@sysucc.org.cn

Abstract

Background: The COVID-19 outbreak has increased challenges associated with health management, especially cancer management. In an effort to provide continuous pharmaceutical care to cancer patients, Sun Yat-sen University Cancer Center (SYSUCC) implemented a remote pharmacy service platform based on its already existing web-based hospital app known as Cloud SYSUCC.

Objective: The aim of this study was to investigate the characteristics, acceptance, and initial impact of the Cloud SYSUCC app during a COVID-19 outbreak in a tertiary cancer hospital in China.

Methods: The total number of online prescriptions and detailed information on the service were obtained during the first 6 months after the remote service platform was successfully set up. The patients' gender, age, residence, primary diagnosis, drug classification, weekly number of prescriptions, and prescribed drugs were analyzed. In addition, a follow-up telephonic survey was conducted to evaluate patients' satisfaction in using the remote prescription service.

Results: A total of 1718 prescriptions, including 2022 drugs for 1212 patients, were delivered to 24 provinces and municipalities directly under the Central Government of China between February 12, 2020, and August 11, 2020. The majority of patients were female (841/1212, 69.39%), and 90.18% (1093/1212) of them were aged 31-70 years old. The top 3 primary diagnoses for which remote medical prescriptions were made included breast cancer (599/1212, 49.42%), liver cancer (249/1212, 20.54%), and thyroid cancer (125/1212, 10.31%). Of the 1718 prescriptions delivered, 1435 (83.5%) were sent to Guangdong Province and 283 (16.5%) were sent to other provinces in China. Of the 2022 drugs delivered, 1012 (50.05%) were hormonal drugs. The general trend in the use of the remote prescription service declined since the 10th week. A follow-up telephonic survey found that 88% (88/100) of the patients were very satisfied, and 12% (12/100) of the patients were somewhat satisfied with the remote pharmacy service platform.

Conclusions: The remote pharmacy platform Cloud SYSUCC is efficient and convenient for providing continuous pharmaceutical care to patients with cancer during the COVID-19 crisis. The widespread use of this platform can help to reduce person-to-person transmission as well as infection risk for these patients. Further efforts are needed to improve the quality and acceptance of the Cloud SYSUCC platform, as well as to regulate and standardize the management of this novel service.

(J Med Internet Res 2021;23(1):e24619) doi:10.2196/24619

KEYWORDS

COVID-19; cancer patients; remote pharmacy; service platform; implementation; oncology; pharmacy; online platform; cancer; health management; app; online hospital; acceptance; impact

Introduction

China and the rest of the world are experiencing an outbreak of the highly infectious, novel COVID-19 [1-3]. To effectively combat the COVID-19 pandemic, Chinese health authorities have effected a series of preventive and control measures since January 21, 2020, including restrictions on people's movement, reduced transportation, entry and exit controls for towns and villages, and isolation requirements for travelers between various parts of the country [4]. However, for patients with cancer, restrictions on movement could result in disease upstaging and have considerable impact on cancer-specific death rates due to delayed diagnoses and the lack of or suboptimal cancer management. Therefore, the management of patients with cancer has become a crucial issue in cities with large-scale outbreaks of COVID-19 [5-7].

Since May 24, 2018, Sun Yat-sen University Cancer Center (SYSUCC) in China has implemented a hospital-based, mobile app called Cloud SYSUCC that enables remote medication consultation between doctors and patients, available both on Android and iOS mobile operating systems. This app includes 2 introductory screens: one for the medical staff (ie, therapeutic interface) and the other for patients (ie, patient interface; see Figure S1 in Multimedia Appendix 1). The therapeutic interface has been designed for doctors to check a patient's medical history, communicate with them, and order examinations; however, it did not provide access to medication prescriptions. For this reason, a multidisciplinary working group comprising senior hospital pharmacists, clinical experts, and information technology engineers was formed on January 23, 2020, to set up a remote pharmacy service platform based on the Cloud SYSUCC app. This platform would enable remote services such as web-based consultation, prescription, dispensation, and home delivery of oral anticancer medications, as well as offer standard administration instructions for the use of these drugs. In this paper, we introduce this remote pharmacy service system and describe its development and implementation in a tertiary cancer hospital in China.

To explore the advantages of the remote pharmacy service system for patients with cancer during public health emergencies, we analyzed the online prescriptions of outpatients at SYSUCC in Guangzhou City, Guangdong Province, China. We collected data from the first 6 months after the remote service platform was successfully set up, and we analyzed the characteristics, user acceptance, and initial impact of this new bundled approach during the COVID-19 outbreak.

Methods

Platform Development

The remote pharmacy service platform provides services to the public via the Cloud SYSUCC app, which has been deployed on a public cloud. A private cloud connects the public cloud with the internal hospital information system using a web services internet tool.

Data Sources

The total number of online prescriptions and their detailed information were obtained automatically from the hospital information system of Cloud SYSUCC. We collected data from February 12, 2020, to August 11, 2020. Patient's gender, age, residence, primary diagnosis, drug classification, weekly number of prescriptions, and prescribed drugs were included in the analysis. Furthermore, we also conducted a telephonic survey to evaluate patients' satisfaction with using the remote prescription service. We used SPSS software (version 16.0; IBM Corp) to summarize and analyze the data.

Results

Real-Time Medication Consultation

We integrated the therapeutic interface of Cloud SYSUCC with the internal hospital information system by using a web services internet tool. Each medical staff, including pharmacists, could access the platform by using their unique staff identification number and password. This integration allows clinical pharmacists to provide real-time patient consultation and check the patient's medical history when necessary (Figure 1). They can also query the patient's consultations status (eg, pending or reviewed), and their observations can be documented in the electronic clinical history.

Figure 1. Screenshot of real-time medication consultation between clinical pharmacists and patients.

3rd March 12:02

Hello, I have been prescribed Regorafenib at the hospital. But I don't know whether I use it everyday or only for a few days?

3rd March 12:21

Hello, you should take Regorafenib 160 mg po on days 1-21, followed by a 7-day rest period within a 4-week schedule. Please take Regorafenib at the same time each day. Swallow tablet whole half an hour after a low-fat breakfast and drink 100~200mL warm water.

What are the specific requirements for a low-fat breakfast ?

A low-fat breakfast allows no more than 600 calories and less than 30% of calories from fat.

3rd March 12:44

Thank you. Can I eat a meat bun, a bowl of oatmeal, and an egg in my breakfast?

I need to confirm whether you add milk in your oatmeal and egg in your meat bun.

3rd March 12:58

Yes, I add milk in my oatmeal.

Eggs may bring you too much fat content in your breakfast. We suggest you to have 100 gram meat bun, 100 gram oatmeal without milk and one-half of an egg for your low-fat breakfast.

Thank you for your helpful and detail suggestion.

Online Prescriptions

Considering patients' safety, effectiveness, and emergency medical needs, the Pharmacy and Therapeutics Committee at SYSUCC has approved 22 oral anticancer drugs and 9 adjuvant medicines; these medications are currently listed in the Cloud SYSUCC app (Table 1). Doctors can prescribe the required drugs through the app in compliance with the following guidelines. First, for prescription of chemotherapeutic drugs, their potential adverse events should be considered and adequate surveillance should be allowed; the duration of these prescriptions should not exceed 14 days. Second, nonchemotherapeutic drugs can be prescribed for a duration not exceeding 30 days. Third, to address the need to balance the safety and feasibility of chemotherapeutic drugs, each patient is allowed 2 remote chemotherapy prescription cycles, after which they are required to visit the hospital for adequate examinations before they can continue using the Cloud SYSUCC app. All prescriptions are pharmaceutically verified by an automated drug rationality review system integrated into our cloud-based hospital information system.

Table 1. Anticancer drugs (n=22) and adjuvant medicines (n=9) listed in the Cloud SYSUCC app.

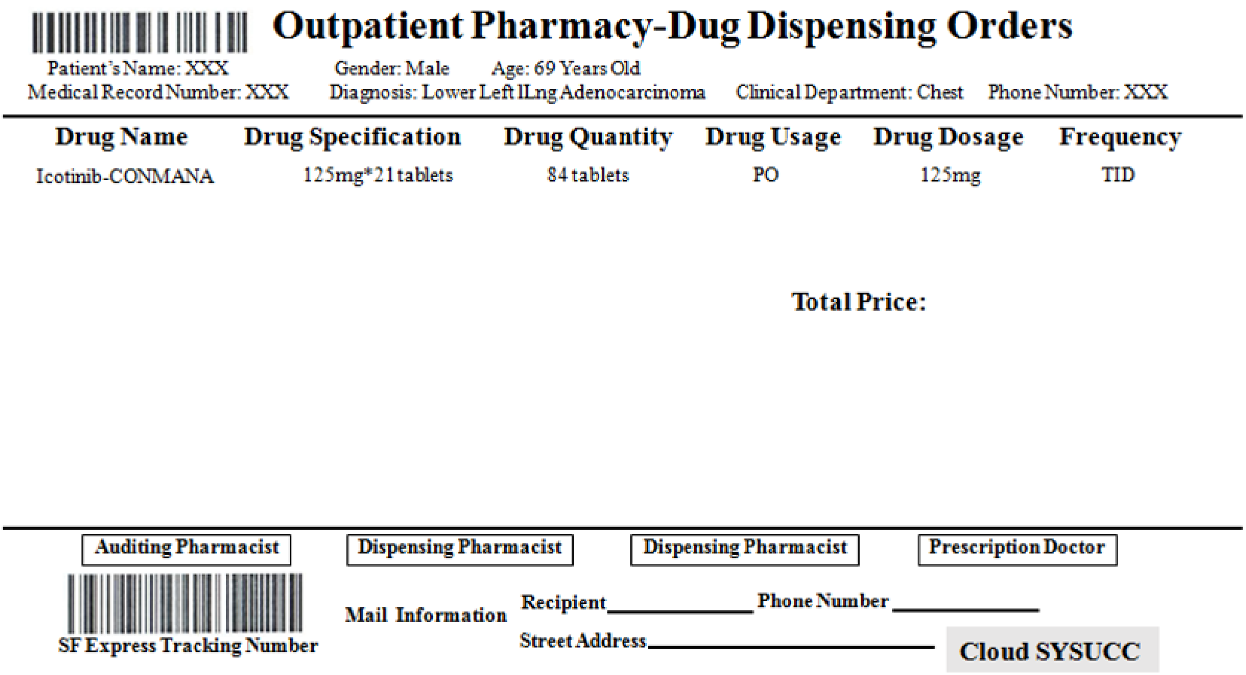
Drug type and name	Dosage form	Dosage strength	Drug quantity	Disease name
Hormonal drugs				
Exemestane	Tablet	25 mg	14	Breast cancer
Anastrozole	Tablet	1 mg	14	Breast cancer
Letrozole	Tablet	2.5 mg	30	Breast cancer
Toremifene	Tablet	60 mg	30	Breast cancer
Tamoxifen	Tablet	10 mg	60	Breast cancer
Medroxyprogesterone	Tablet	500 mg	30	Breast cancer
Megestrol acetate	Tablet	80 mg	24	Breast cancer, endometrial cancer
Bicalutamide	Tablet	50 mg	28	Prostate cancer
Levothyroxine sodium	Tablet	50 µg	100	Thyroid cancer
Calcium preparations				
Calcium carbonate	Tablet	1.5 g	30	Osteoporosis
Alfacalcidol	Capsule	1 µg	10	
Calcitriol	Capsule	0.25 µg	10	
Calcium supplement with vitamin D	Tablet	750 mg	60	
Antiviral drugs				
Entecavir	Tablet	0.5 mg	7	Type B viral hepatitis
Tenofovir	Tablet	0.3 g	30	
Telbivudine	Tablet	600 mg	7	
Lamivudine	Tablet	0.1 g	14	
Adefovir dipivoxil	Tablet	10 mg	14	
Alkylating drugs				
Cyclophosphamide	Tablet	50 mg	24	Hodgkin disease, acute myeloid leukemia, breast cancer
Temozolomide	Capsule	20 mg	5	Glioblastoma multiforme
Temozolomide	Capsule	100 mg	5	Anaplastic astrocytoma
Antimetabolite drugs				
Capecitabine	Tablet	500 mg	12	Colorectal cancer
Tegafur	Tablet	50 mg	100	Gastric cancer, colorectal cancer, breast cancer
Tegafur/Uracil Monopotassium	Tablet	162 mg	20	Gastric cancer, colorectal cancer, breast cancer
Tegafur/Gimeracil/Oteracil Monopotassium	Capsule	20 mg	140	Gastric cancer
Angiogenesis inhibitors				
Thalidomide	Capsule	25 mg	48	Multiple myeloma
Kinase inhibitors				
Gefitinib	Tablet	250 mg	10	Lung cancer
Imatinib	Tablet	100 mg	60	Gastrointestinal stromal tumor, lymphoblastic leukemia

Drug type and name	Dosage form	Dosage strength	Drug quantity	Disease name
Icotinib	Tablet	125 mg	21	Lung cancer
Erlotinib	Tablet	100 mg	30	
		150 mg	7	
Apatinib	Tablet	250 mg	10	Gastric adenocarcinoma, gastroesophageal conjunctive adenocarcinoma
Afatinib	Tablet	30 mg	7	Lung cancer
		40 mg	7	

Drug Dispensation and Home Delivery Service

Once an online prescription is generated from the Cloud SYSUCC app, an alert system linked to the SYSUCC drug dispensing system is initiated. Then, the frontline pharmacists can print the drug dispensing orders with an automatically generated delivery service tracking number (SF Express Group Co., Ltd.; Figure 2). All drug orders are double-checked before being transferred to SF Express, which provides a quick, cost-effective, and meticulous Mainland China Express Service to ensure safe and efficient delivery. Patients can view live updates by using the tracking information of the drug orders they have placed via the Cloud SYSUCC app. To ensure the drugs are delivered correctly to patients, SF Express delivery agents verify the identity of the patients before handing out the packages. If the delivery of packages takes more than 3 days, the patients can simply request a refund for the express order charges.

Figure 2. Example of a drug dispensing order generated using the Cloud SYSUCC app.



Outpatient Pharmacy-Dug Dispensing Orders

Patient's Name: XXX Gender: Male Age: 69 Years Old
Medical Record Number: XXX Diagnosis: Lower Left lLng Adenocarcinoma Clinical Department: Chest Phone Number: XXX

Drug Name	Drug Specification	Drug Quantity	Drug Usage	Drug Dosage	Frequency
Icotinib-CONMANA	125mg*21tablets	84 tablets	PO	125mg	TID

Total Price:

Auditing Pharmacist
Dispensing Pharmacist
Dispensing Pharmacist
Prescription Doctor

SF Express Tracking Number

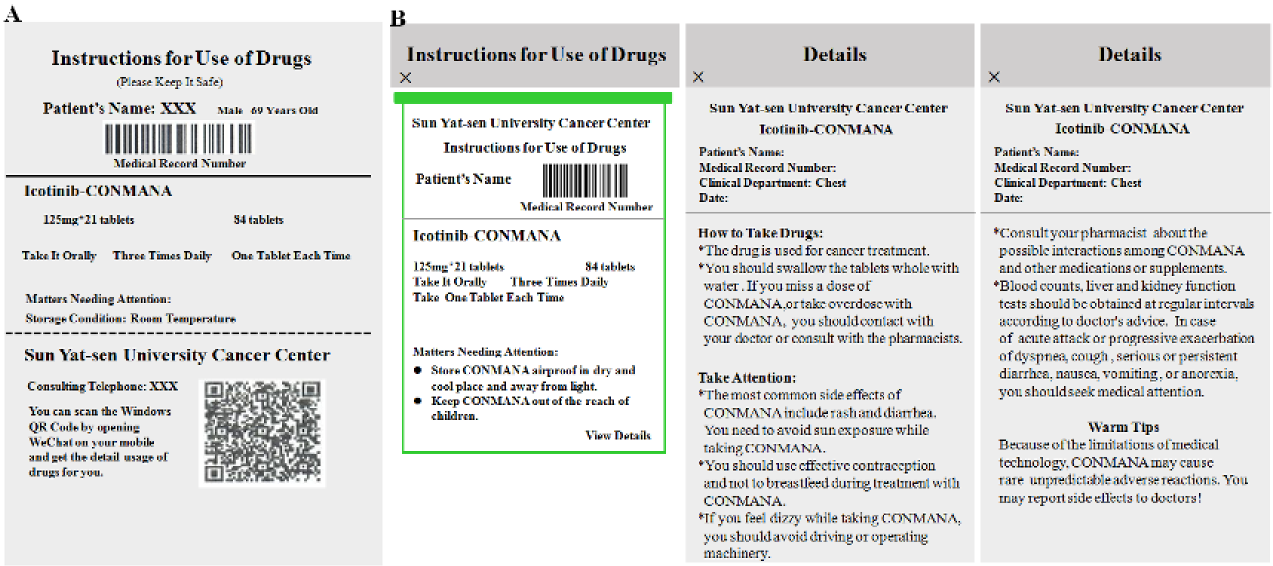
Mail Information
Recipient__________ Phone Number__________
Street Address____________________

Cloud SYSUCC

Standard Administration Instructions

To ensure safe handling of oral anticancer drugs by patients at home, clinical pharmacists have undertaken the following essential measures: (1) prescription directions are printed on each drug box; (2) detailed directions for the usage of drugs are made available on the patient's interface once the drugs are ordered (Figure 3A); and (3) patients can use their mobile phone's camera to scan a QR (quick response) code that will prompt a WeChat (Tencent Tech Shenzhen Co., Ltd.) window on their phone, wherein also detailed drug usage directions will be provided. These measures will ensure patients have easy access to the detailed instructions whenever they need (Figure 3B). If patients have any concerns regarding the drugs, they can consult with their treating doctors or clinical pharmacists using the Cloud SYSUCC app.

Figure 3. Screenshots of drug instructions shared with patients ordering drugs via the Cloud SYSUCC app: (A) printed version and (B) electronic version.



Instructions for Use of Drugs

(Please Keep It Safe)

Patient's Name: XXX Male 69 Years Old

Medical Record Number

Icotinib-CONMANA

125mg*21 tablets 84 tablets

Take It Orally Three Times Daily One Tablet Each Time

Matters Needing Attention:

Storage Condition: Room Temperature

Sun Yat-sen University Cancer Center

Consulting Telephone: XXX

You can scan the Windows QR Code by opening WeChat on your mobile and get the detail usage of drugs for you.

Instructions for Use of Drugs

Sun Yat-sen University Cancer Center

Instructions for Use of Drugs

Patient's Name

Medical Record Number

Icotinib-CONMANA

125mg*21 tablets 84 tablets

Take It Orally Three Times Daily

Take One Tablet Each Time

Matters Needing Attention:

- Store CONMANA airproof in dry and cool place and away from light.
- Keep CONMANA out of the reach of children.

Details

Sun Yat-sen University Cancer Center

Icotinib-CONMANA

Patient's Name:
Medical Record Number:
Clinical Department: Chest
Date:

How to Take Drugs:

*The drug is used for cancer treatment.

*You should swallow the tablets whole with water . If you miss a dose of CONMANA,or take overdose with CONMANA, you should contact with your doctor or consult with the pharmacists.

Take Attention:

*The most common side effects of CONMANA include rash and diarrhea. You need to avoid sun exposure while taking CONMANA.

*You should use effective contraception and not to breastfeed during treatment with CONMANA.

*If you feel dizzy while taking CONMANA, you should avoid driving or operating machinery.

Details

Sun Yat-sen University Cancer Center

Icotinib CONMANA

Patient's Name:
Medical Record Number:
Clinical Department: Chest
Date:

*Consult your pharmacist about the possible interactions among CONMANA and other medications or supplements.

*Blood counts, liver and kidney function tests should be obtained at regular intervals according to doctor's advice. In case of acute attack or progressive exacerbation of dyspnea, cough , serious or persistent diarrhea, nausea, vomiting , or anorexia, you should seek medical attention.

Warm Tips

Because of the limitations of medical technology, CONMANA may cause rare unpredictable adverse reactions. You may report side effects to doctors!

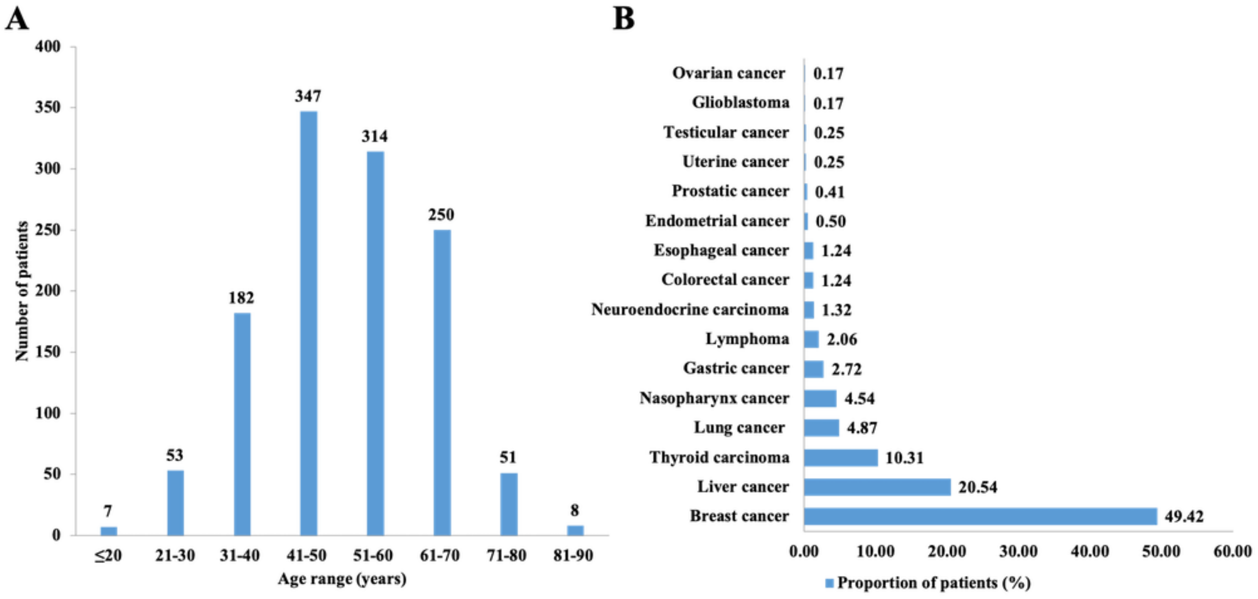
Patient Demographics and Distribution of Online Prescriptions

During the study period (February 12 to August 11, 2020), a total of 1718 prescriptions, including 2022 drugs for 1212 patients, were delivered to 24 provinces and municipalities directly under the Central Government of China. The patient demographics are shown in Table 2. The majority of patients were female (841/1212, 69.39%). The mean age of patients was 51.26 (SD 12.29) years (age range: 12-87), and 90.18% (1093/1212) of the patients were aged between 31 and 70 years old (Figure 4A). The distribution of patients based on their primary diagnosis showed that breast cancer, liver cancer, and thyroid cancer were the 3 most common types of cancers treated (49.42%, 20.54%, and 10.31%, respectively; Figure 4 B) with prescriptions using this remote platform.

Table 2. Characteristics of patients who received online prescriptions from the Cloud SYSUCC app (N=1212).

Characteristic	Value
Gender, n (%)	
Male	371 (30.61)
Female	841 (69.39)
Age (years), median (range)	51 (12-87)
Age group (years), n (%)	
0-10	0 (0.00)
10-20	7 (0.58)
21-30	53 (4.37)
31-40	182 (15.02)
41-50	347 (28.63)
51-60	314 (25.91)
61-70	250 (20.63)
71-80	51 (4.21)
81-90	8 (0.66)
Primary diagnosis, n (%)	
Breast cancer	599 (49.42)
Liver cancer	249 (20.54)
Thyroid carcinoma	125 (10.31)
Lung cancer	59 (4.87)
Nasopharyngeal cancer	55 (4.54)
Gastric cancer	33 (2.72)
Lymphoma	25 (2.06)
Neuroendocrine carcinoma	16 (1.32)
Colorectal cancer	15 (1.24)
Esophageal cancer	15 (1.24)
Endometrial cancer	6 (0.50)
Prostatic cancer	5 (0.41)
Uterine cancer	3 (0.25)
Testicular cancer	3 (0.25)
Glioblastoma	2 (0.17)
Ovarian cancer	2 (0.17)

Figure 4. Distribution of patients by age (A) and primary diagnosis (B) using data from the Cloud SYSUCC app (N=1212).



Of the 1718 delivered prescriptions, 1435 (83.5%) prescriptions were delivered to Guangdong Province, including 437 (25.4%) in Guangzhou city, and the remaining 283 (16.5%) prescriptions were delivered to other provinces in China (Figure 5). The top 5 provinces for out-of-province prescription deliveries were Hunan, Guangxi, Jiangxi, Hainan, and Fujian, all of which are located in southern China (Figure 5).

Figure 5. Regional distribution of prescription deliveries via the Cloud SYSUCC app (N=1718).

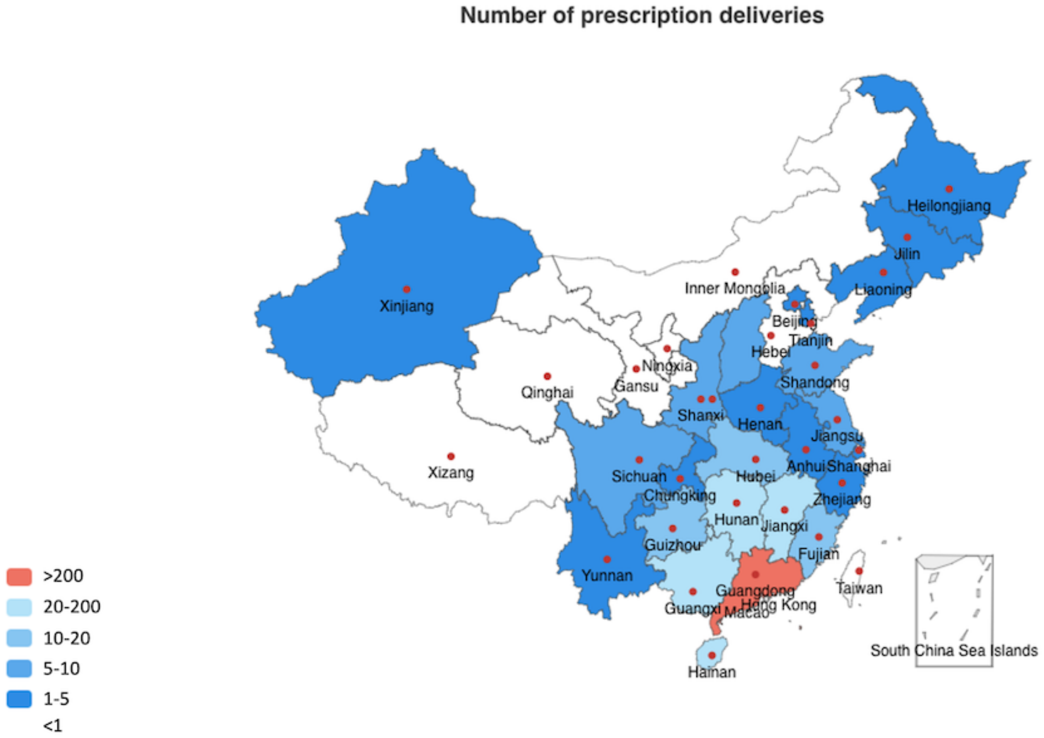
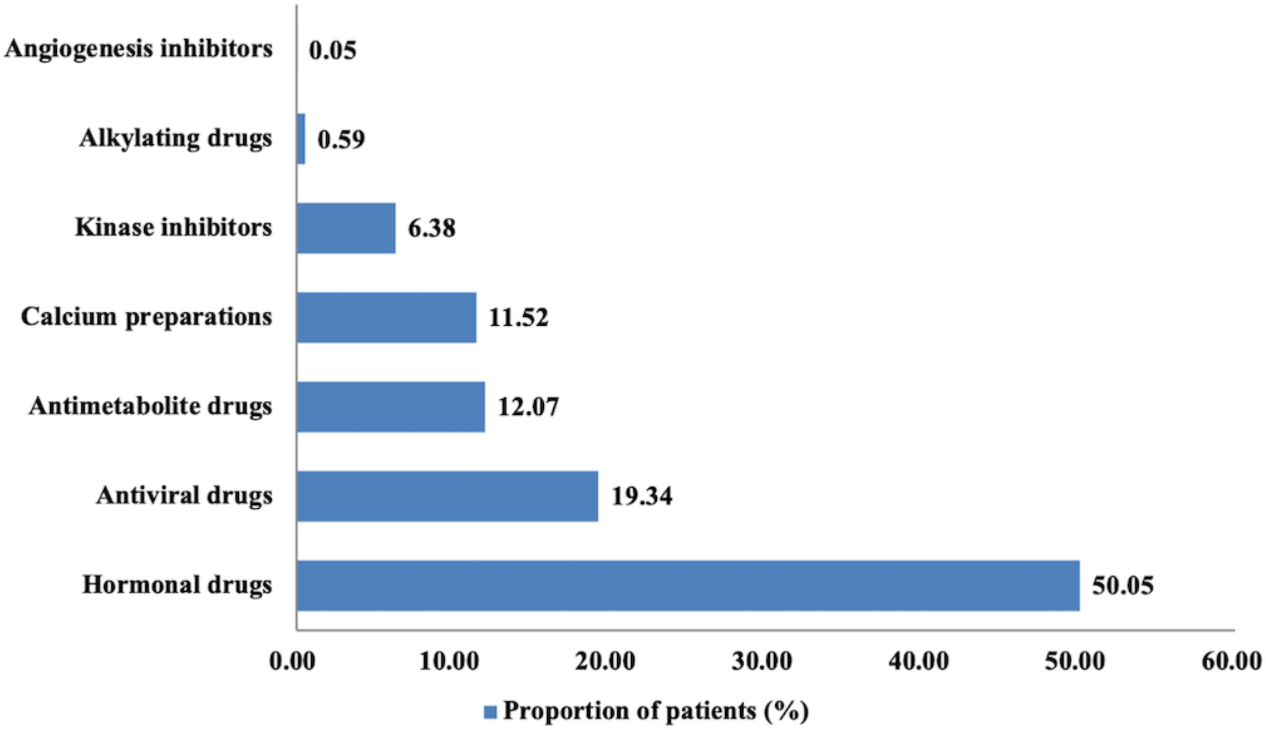


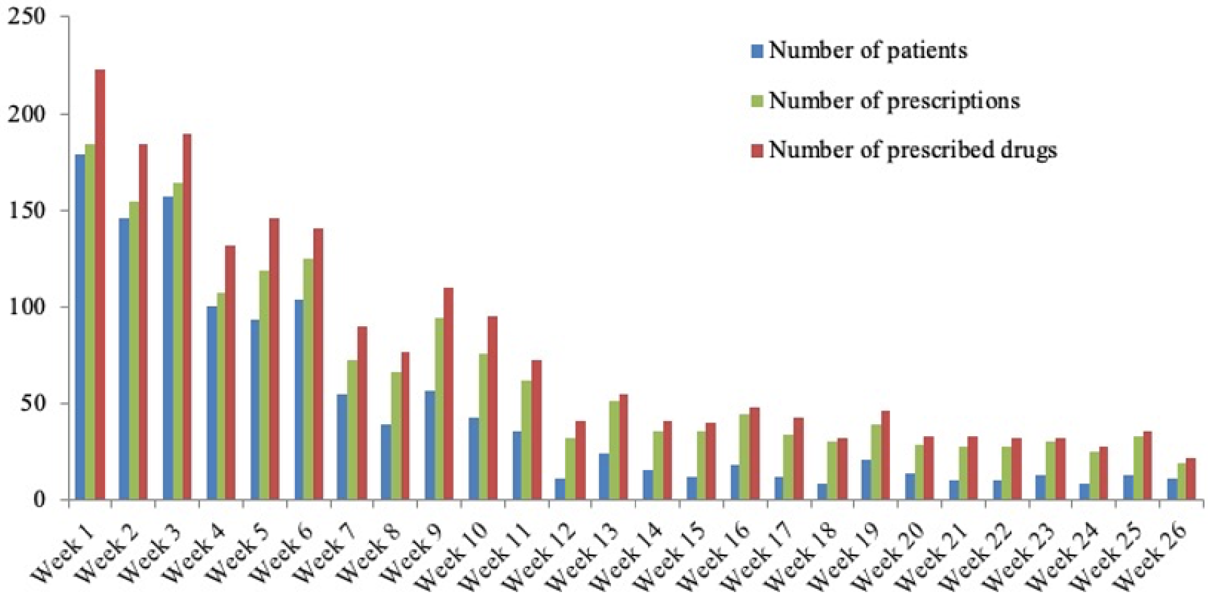
Figure 6 shows the distribution of drugs delivered (mean 1.19, SD 0.46; range: 1-6) to the 1212 patients based on drug classification. Among the 2022 drugs delivered, 1012 (50.05%) were hormonal drugs. Most hormonal drugs were used to treat breast cancer and thyroid cancer, which was consistent with the distribution of the patients by primary diagnosis.

Figure 6. Distribution of online prescriptions by classification of drugs ordered via the Cloud SYSUCC app (N=2022).



The weekly number of prescriptions, prescribed drugs, and patients pertaining to these prescriptions are summarized in Figure 7. As the COVID-19 outbreak was brought under control in China, the general trend in the use of the remote prescription service considerably declined. The total number of patients who used the remote prescription service in the first 9 weeks of the outbreak was 930; this number was 3 times higher than the number of patients who used the service in the following 17 weeks.

Figure 7. Weekly number of patients who used the remote prescription system, medication prescriptions ordered, and prescribed drugs delivered via the Cloud SYSUCC app since February 12, 2020.



Telephonic Patient Satisfaction Survey Evaluating the Remote Prescription Service

A small-scale telephonic survey was conducted to obtain patients' views about the remote prescription service. A total of 100 patients were randomly selected from the total sample of 1212 patients over a 6-month period. These patients were contacted 7 days after their prescriptions were delivered. This time lag between prescription delivery and inquiry was intended to allow patients sufficient time to receive the drugs and reflect on their experiences. The survey contained only 1 simple question: “What do you think about the remote prescription service?” Patients were asked to choose from the following 5 multiple choice answers: (1) satisfied, (2) somewhat satisfied, (3) neutral, (4) somewhat dissatisfied, and (5) very dissatisfied. The survey question was very specific to the remote prescription service so that patients were clear about what was being asked and could accurately express their actual experience. If a selected patient declined to participate in the survey, another patient would be chosen to substitute him or her. The use of randomization to select the survey sample was to ensure a general representation of the population of study patients using the remote prescription service. Thus, bias in the patient sample, such as age, gender, and social class was reduced. We believe that this randomization of the sample enhanced the external validity of the survey results. The telephonic survey showed that 88% (88/100) of the participants were satisfied, whereas 12% (12/100) of them were somewhat satisfied. None of the patients selected a neutral, somewhat dissatisfied, or very dissatisfied response. Overall, these findings suggest a high level of user satisfaction with our remote pharmacy service platform.

Discussion

Principal Findings

We conducted a pilot evaluation of the remote pharmacy service platform in a tertiary cancer hospital in Guangzhou, China, during the first 6 months after work resumption. By September 20, 2020, the rapid spread of COVID-19 had resulted in 30,675,675 cases and 954,417 deaths worldwide. Traditional hospitals required patients to visit the hospitals for obtaining medications, which could potentially cause more infections and lead to severe health deterioration during an ongoing epidemic, particularly for patients with cancer who have a suppressed immune system. Studies have shown that patients with cancer have a higher risk of severe events than do other patients [8,9]. Moreover, a delayed diagnosis and suboptimal cancer management due to the pandemic could be life-threatening for these patients [5,6]. Therefore, access to timely medications has become a crucial issue during large-scale outbreaks of COVID-19. Our remote pharmacy service platform can prove beneficial for such use, with a good satisfactory response from patients with cancer. To our knowledge, this remote pharmacy service platform for patients with cancer is the first such information system introduced worldwide, comprising a collaborative mechanism that integrates the Cloud SYSUCC app with the internal hospital information system and facilitates remote consultation service by using a standardized format for ethical drug prescription and standard administration instructions.

From February 12, 2020, to August 11, 2020, a total of 60,968 users used the Cloud SYSUCC app, an average of 500 users per day. This resulted in 1718 prescriptions and US $3054 million in drugs fees. Most of the medical consultations were for prehospital services such as cancer diagnosis, cancer type detection, and cancer treatment. The number of online prescriptions recorded increased from the first week investigated to the 6th week (Figure 7), which indicates the acceptance of Cloud SYSUCC app among the users. The decline in prescription numbers since the 7th week might be the result of 2 factors, namely, the revision in the Cloud SYSUCC registration fee from US $7 to US $70 per user and the fact that the COVID-19 outbreak was brought under control.

Most patients using the prescription service were female with a breast cancer diagnosis, aged between 31 and 70 years, and had an existing prescription for hormonal drugs. Since hormone-receptor-positive, early-stage breast cancer is a chronic disease, continuous daily dosing of hormone therapy after surgery is very important in these cases. However, at the same time, hormone therapy may cause serious adverse reactions, not only affecting the patient’s quality of life but also eventually leading to therapeutic resistance. Therefore, timely evaluation and adaptation of treatment is needed for these patients. Hence, the Pharmacy and Therapeutics Committee at SYSUCC has authorized doctors to prescribe the required drugs through the Cloud SYSUCC app provided that the prescriptions are compliant with specific guidelines to balance safety and feasibility of the use of anticancer drugs. Moreover, exclusive medical education campaigns are carried out for older patients who were recommended to stay at home and avoid contact with other people for an extended period during the COVID-19 outbreak [10]. Our results showed that only 4.87% of the patients requiring online prescriptions were aged ≥70 years (Table 2). This may be due to the differences in public acceptance and uncertainty of physical conditions for older-aged populations.

Future Prospects

Thus far, the COVID-19 outbreak in China has been under control, and there is no rapid increase in the number of new cases. However, the possibility of a renewed spike in COVID-19 cases in individual regions cannot be ruled out. Therefore, more effort is needed to make better use of the Cloud SYSUCC–based remote pharmacy service during and after the current pandemic. Simple and clear instructions are necessary to improve acceptance of this platform by older patients. Improving financial support, such as by reducing the Cloud SYSUCC registration fee, can also promote increased adoption of the platform by the public. Moreover, timely web-based pharmaceutical care interventions and patient monitoring are required to ensure the safety of patients with cancer.

Conclusions

Currently, the world is fighting against the COVID-19 pandemic, which has affected over 200 countries and regions. The findings of this study suggest that the Cloud SYSUCC remote pharmacy platform is efficient and convenient for enabling continuous pharmaceutical care to patients with cancer. The widespread use of this platform can help reduce both person-to-person transmission of COVID-19 and the infection risk for patients with cancer.

Acknowledgments

We thank Prof Hong-Sheng Wang of the School of Pharmaceutical Sciences, SYSUCC, for their assistance with the revision of this manuscript. This research was supported by the National Natural Science Foundation of China (Grant No. 81672608), and the outstanding young talent promotion scheme for the construction of high-level hospitals in Guangdong Province (No. PT19210101).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the home page of the Cloud SYSUCC app: (A) therapeutic interface and (B) patient interface.

[DOCX File , 204 KB - jmir_v23i1e24619_app1.docx]

References

1. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020 Feb 22;395(10224):565-574 [FREE Full text] [doi: 10.1016/S0140-6736(20)30251-8] [Medline: 32007145]
2. Cao Y, Liu X, Xiong L, Cai K. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2: A systematic review and meta-analysis. J Med Virol 2020 Sep;92(9):1449-1459 [FREE Full text] [doi: 10.1002/jmv.25822] [Medline: 32242947]
3. Zhai P, Ding Y, Wu X, Long J, Zhong Y, Li Y. The epidemiology, diagnosis and treatment of COVID-19. Int J Antimicrob Agents 2020 May;55(5):105955 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105955] [Medline: 32234468]
4. China takes top-level preventive, control measures against 2019-nCoV pneumonia. English.gov.cn - The State Council, The People's Republic of China. 2020 Jan 21. URL: http://english.www.gov.cn/statecouncil/ministries/202001/21/content_WS5e265d25c6d0db64b784ce80.html [accessed 2021-01-06]
5. Wang H, Zhang L. Risk of COVID-19 for patients with cancer. Lancet Oncol 2020 Apr;21(4):e181 [FREE Full text] [doi: 10.1016/S1470-2045(20)30149-2] [Medline: 32142621]
6. Al-Shamsi HO, Alhazzani W, Alhuraiji A, Coomes EA, Chemaly RF, Almuhanna M, et al. A practical approach to the management of cancer patients during the novel coronavirus disease 2019 (COVID-19) pandemic: an international collaborative group. Oncologist 2020 Jun;25(6):e936-e945 [FREE Full text] [doi: 10.1634/theoncologist.2020-0213] [Medline: 32243668]
7. Ueda M, Martins R, Hendrie PC, McDonnell T, Crews JR, Wong TL, et al. Managing cancer care during the COVID-19 pandemic: agility and collaboration toward a common goal. J Natl Compr Canc Netw 2020 Mar 20:1-4. [doi: 10.6004/jnccn.2020.7560] [Medline: 32197238]
8. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. The Lancet Oncology 2020 Mar;21(3):335-337. [doi: 10.1016/s1470-2045(20)30096-6]
9. Dai M, Liu D, Liu M, Zhou F, Li G, Chen Z, et al. Patients with cancer appear more vulnerable to SARS-CoV-2: a multicenter study during the COVID-19 outbreak. Cancer Discov 2020 Jun;10(6):783-791 [FREE Full text] [doi: 10.1158/2159-8290.CD-20-0422] [Medline: 32345594]
10. Brooke J, Jackson D. Older people and COVID-19: isolation, risk and ageism. J Clin Nurs 2020 Jul;29(13-14):2044-2046. [doi: 10.1111/jocn.15274] [Medline: 32239784]

Abbreviations

SYSUCC: Sun Yat-sen University Cancer Center

Edited by G Eysenbach; submitted 28.09.20; peer-reviewed by E Tashkandi, R Vilela; comments to author 27.11.20; revised version received 07.12.20; accepted 24.12.20; published 21.01.21.

Please cite as:
Chen ZJ, Liang WT, Liu Q, He R, Chen QC, Li QF, Zhang Y, Du XD, Pan Y, Liu S, Li XY, Wei X, Huang H, Huang HB, Liu T
Use of a Remote Oncology Pharmacy Service Platform for Patients With Cancer During the COVID-19 Pandemic: Implementation and User Acceptance Evaluation
J Med Internet Res 2021;23(1):e24619
URL: http://www.jmir.org/2021/1/e24619/
doi:10.2196/24619
PMID:33395398

Viewpoint

Patient Care During the COVID-19 Pandemic: Use of Virtual Care

Andy Wong[1], MSc, MD; Rashaad Bhyat[2], MB, BCh, CCFP; Siddhartha Srivastava[3], MD; Lysa Boissé Lomax[1], MSc, MD; Ramana Appireddy[1], MBBS, DM, MSc

[1]Division of Neurology, Queen's University, Kingston, ON, Canada
[2]Canada Health Infoway, Toronto, ON, Canada
[3]Department of Medicine, Queen's University, Kingston, ON, Canada

Corresponding Author:
Ramana Appireddy, MBBS, DM, MSc
Division of Neurology
Queen's University
76 Stuart Street
Connell 7
Kingston, ON, K7L 2V7
Canada
Phone: 1 6135482472
Email: mrra@queensu.ca

Abstract

Virtual care, the use of videoconferencing technology to connect with patients, has become critical in providing continuing care for patients during the current COVID-19 pandemic. Virtual care has now been adopted by health care providers across the spectrum, including physicians, residents, nurse practitioners, nurses, and allied health care professionals. Virtual care is novel and nuanced compared to in-person care. Most of the health care providers who are delivering or expected to deliver virtual care have little to no prior experience with it. The nuances of virtual care involve regulatory standards, platforms, technology and troubleshooting, patient selection, etiquette, and workflow, all of which comprise critical points in the provision of health care. It is important to consistently deliver high-quality, equitable, and professional virtual care to inspire patients with the trust they need to continue follow-up of their care in these difficult times. We have been adopting virtual care in our clinical practice for over two years. In partnership with Canada Health Infoway, we have assembled a primer for virtual care that can serve as a guide for any health care provider in Canada and globally, with the goal of providing seamless transitions between in-person and virtual care.

(J Med Internet Res 2021;23(1):e20621) doi:10.2196/20621

KEYWORDS

virtual care; teleneurology; telemedicine; medical informatics; internet; patient-physician relationship; email; digital health

Introduction to Virtual Care

As part of the response to the COVID-19 pandemic, health care organizations across Canada have cancelled elective and nonurgent clinics as a measure to reduce the risk of exposing patients to COVID-19 [1]. To provide safe, timely, and accessible ambulatory care, health care providers have adopted virtual care in Canada and globally [2-7]. Virtual care can be defined as any interaction between patients and members of their circle of care that occurs remotely, using any form of communication or information technology, to facilitate or maximize the quality and effectiveness of patient care [8-11].

Role of Virtual Care in Addressing Barriers to Health Care

Virtual care is novel and nuanced compared to in-person care. Virtual Care has significant benefits to patients, health care systems and society at large by offering patient-centered care [12-16]. Virtual care also helps increase health care capacity by optimal use of health care professionals' time, infrastructure, and reduced per capita health care cost [13]. Virtual care is more relevant during the COVID-19 pandemic because it provides access to medical care that is timely, convenient, efficient, and safe with reduced risk of transmission [17]. Despite all these advantages of virtual care, it is important to ensure that virtual care continues to provide equitable access to health care. Virtual care has the potential to remove some of the current barriers to health care with regard to availability, accessibility and affordability [13,18-20]. Geographical distance, travel burden and out-of-pocket expenses are some of the largest barriers to ambulatory health care access, and virtual care has great potential to address these [2,21]. However, sociodemographic characteristics such as age, sex, gender, level of education, and English proficiency, as well as socioeconomic status indicators such as race and ethnicity, can impact one’s ability to access and use the technology needed for participating in virtual care [22-24]. Virtual care has the potential to provide access to people who live in more rural areas, who tend to have lower socioeconomic status, are older, and have lower levels of education, but only if the patient can access video and audio equipment, spend the time to use and troubleshoot the equipment, and learn to use programs and hardware they may have never used before. There is a need for engagement and commitment from all stakeholders in the health care system to ensure that virtual care can successfully provide equitable access, closing the gap between rural and urban health care [25].

The Division of Neurology at Queen's University/Kingston Health Sciences Center has been an early adopter of virtual care for ambulatory services since 2016 for stroke, epilepsy, and sleep medicine, and it has an active virtual care research program [13,26]. Leveraging our long experience with virtual care, the Southeastern Ontario Academic Medical Organization and Kingston Health Sciences Center have enabled a rapid rollout of virtual care services for its 450 physicians, 575 medical residents, and 160 allied health care professionals over two weeks using the Reacts platform [27]. Many health care providers (physicians, residents, medical students, nurses, and allied health care professionals) are not familiar with using virtual care and have identified virtual care and digital health technologies as areas of emerging educational and training focus [9-11,27,28]. Here, we present a comprehensive, practical guide for any health care provider wishing to provide virtual care.

Guide to Virtual Care

Licensing and Regulatory Standards

At this time, no rigorous standalone virtual care guidelines have been established by licensing bodies in Canada [9]. Some guidelines have been temporarily developed as a response to the COVID-19 crisis and refer to pre-existing telemedicine guidelines that may require rejuvenation [29,30]. The Canadian Medical Association is providing guidance through the Virtual Care Playbook [31]. Provincial and territorial professional organizations for physicians have released some local guidance on virtual care [3,32-35]. In addition, an excellent pan-Canadian list of virtual care resources is available through the Royal College of Physicians and Surgeons [36].

Further information on medicolegal liability can be obtained from the Canadian Medical Protective Association (CMPA) [37,38]. Nonphysician health care providers (nurses, nurse practitioners, and other allied health care staff) are advised to refer to their own professional and licensing bodies for further guidance. It may be reasonable to refer to the above resources for general guidance on providing virtual care.

Types of Virtual Care and Platforms to Use

Secure messaging, secure email, and secure video conferencing are some of the most used virtual care modalities. These services can be delivered on multiple platforms and devices. The platforms are broadly divided into regulated and unregulated categories. Regulated platforms are those that comply with Canada's federal Personal Information Protection and Electronics Document Act (PIPEDA), as well as provincial and territorial privacy laws that apply to health or medical records, such as Ontario's Personal Health Information Protection Act (PHIPA). A comprehensive list of regulated platforms in Canada is available [3,36]. Unregulated platforms include those that do not meet the federal and provincial/territorial privacy standards (FaceTime, WhatsApp, Google Duo, etc).

For the virtual care provider, the choice of platform may depend on the features offered, such as secure messaging, video conferencing, patient portals, integration into electronic health records (EHRs), and integration with other remote monitoring devices and applications. Before finalizing a decision on a platform, we recommend that practitioners seek clearance from their own organization's privacy and security officer if available. Each organization and each individual provider will have a unique set of circumstances, resources, and abilities. It is essential for each provider to develop familiarity with the platform selected by their organization. The platforms currently used at Queen's University are the Ontario Telemedicine Network and Reacts [27].

Digital Privacy

It is important to ensure that the appropriate precautions are taken by the physician and the patient to ensure the privacy of health care information [39]. For additional information on privacy, security, and data stewardship considerations relating to digital health technologies in the outpatient setting, clinicians should consider guidance from the College of Family Physicians of Canada and the CMPA or from local or regional professional associations [40,41].

Equipment

The equipment (or hardware) required for delivering virtual care is widely available and includes smartphones, tablets, laptop computers, and desktop computers. Ideally, all these devices should have built-in or peripheral hardware for video conferencing (web camera, speaker, and microphone). The network (bandwidth) requirements must be confirmed before use, although most of these platforms enable the quality of the video conference to be adjusted to low bandwidth. In general, the faster and more reliable the internet connection, the better the experience will be for the provider and the patient.

The choice of the device used by the provider depends on multiple factors, including the physical location of the provider at the time of the virtual video visit. Although virtual visits can be conducted from any internet-enabled device, we suggest using a laptop or a desktop computer with a built-in camera and speaker. These devices typically provide the highest quality internet connection, a large screen, and a better speaker and microphone setup. However, this is highly dependent on the technology available to the patient, as modern-day tablets and smartphones have high-quality built-in cameras and speakers. Perhaps most importantly, the patient and provider should be familiar with their respective devices in case troubleshooting is required.

Patient Selection for Virtual Care

It is essential to identify when virtual care is most appropriate for patients. Virtual care is an effective alternative to in-person care for many but not all medical conditions and clinical scenarios. Some general guidelines on patient selection and screening are provided in table 1 [13]. Given the nuances of clinical medicine, with its multiple specialties and subspecialties, health care providers should use professional judgement when deciding which patients, medical conditions, and types of clinical encounters are most suitable for virtual visits [29,38]. As a general guide, virtual care is ideal for nonacute or routine clinical activities, such as reviewing test results and symptom follow-up.

Table 1. Guide to virtual care channel selection.

Virtual care channel	Ideal patient	Ideal conditions	Nonideal conditions
Video	• Has access to required technology and the internet • Is comfortable with using video technology • Has no physical/sensory/cognitive disabilities that would limit videoconference use (unless supported by family and/or friends)	• All clinical conditions or encounters that do not need hands-on physical examination • Follow-up of test results • Counselling • Discussing treatment options • Discussing prognosis • Follow-up of symptom severity	• Workup of complex physical conditions needing physical hands-on examination for diagnosis and therapeutic planning
Secure messaging/email/chat	• Has access to email/chat • Is comfortable with using the modality • Has physical or sensory disabilities that make video conferencing/in-person/telephone difficult • Has a busy work schedule	• Follow-up of routine test results that are normal and do not require intervention • Answering patient questions about medications • Conditions that need very frequent ongoing follow-up • Communication regarding upcoming test results	• Initial encounters with patients • Workup of new symptoms, especially complex ones • Discussing complex diagnostic/therapeutic plans • Initiating new treatments that are complex • Breaking bad news • Providing prognostic information on major medical illness
Telephone	• Has barriers to technology/internet access • Lacks comfort using video technology	• Follow-up of test results • Follow-up of symptom severity • Follow-up of treatment response • Counselling • Discussing treatment options • Treatment titration • Discussion of prognosis	• Initial encounters with patients • Workup of new symptoms, especially complex ones

Virtual Care Workflow: Registration, Appointment, Scheduling, and Documentation

It is mandatory in all jurisdictions to document all health care encounters as part of the patient's health care record. Provincial and territorial medical regulatory authorities and the CMPA consider appropriate documentation to be an essential communication skill and a core component of physicians' professional best practice.

Although some virtual care platforms integrate the video visit into the provider's EHR, patient portal, or hospital information system, this practice is not yet commonplace. Thus, two parallel workflows often coexist for the provider: the first for the virtual care platform, and the second for the provider's medical record system.

This “double workflow” ensures that the virtual care encounter details (registration, clinical documentation, follow-up plans, planned investigations, and therapeutic plans) are captured in the patient's health record.

Consent

It is necessary practice to discuss consent with the patient at the time they sign up for virtual care, covering subjects such as the inherent risks related to privacy and security as well as the limitations of virtual care [32]. There are two routes by which patient information could leave the circle of care. One route is the “real world.” Nonprivate, nonstandard locations on both the patient’s and clinician’s side are now used; this increases the risk that another party will hear the discussion, whether accidentally or intentionally. The precautions described within this article will help to reduce this risk. The other route is the “virtual world,” where there is an inherent risk in the web-based connection between the two parties. Internet and virtual care security have come a long way from their beginnings, and the weak link in information security often lies in the user’s hands. All our connections from the clinician’s side necessitate standard password complexity and 2-factor authentication, including on personal devices that may be used remotely. We also only use software designed and approved for the exchange of personal health information.

The limitations of virtual care should be made clear to the patient. It should be clear that when the clinician cannot see the patient physically, there is inherent risk of missing physical findings. This is only exacerbated by the potential for unstable network connections or poor quality/malfunctioning equipment. It should also be made clear to patients that with this in mind, should there be a requirement for in-person assessment, there may a delay relative to the traditional method of seeing the patient in-person. The option of seeing a clinician in person should always be given to the patient and should always be suggested by the clinician if they feel that the virtual visit was insufficient for the patient's care or may even have caused harm.

Generic consent forms are readily available on the web through many of the resources referenced in this paper [31,37,39], although consent is often obtained verbally given the context and must be documented in the subsequent notes. Although regulated health care software applications require the patient to consent at the time of account signup, informed consent should be obtained from the patient or substitute decision maker before beginning any virtual care interaction.

Patient Setup and Education

It is crucial to gauge the degree of a patient's or caregiver's comfort level with using technology before proposing the use of virtual care. Some health conditions can affect motor strength, coordination, language, vision, and cognition. All these challenges can adversely affect a patient's ability to use personal internet-enabled devices and participate in virtual care, even if the patient is proficient with technology. In addition, it is important to recognize that not all patients will have easy access to the ideal technology for virtual care, including video capability and a sufficiently fast internet connection. We have found that modern mobile phones are capable and ubiquitous but may have poor video and/or audio quality depending on the internet connection. Some patients have chosen to use audio-only telephone visits instead. As we transition to virtual care, we need to be keenly aware of patient-readiness for this transition. Although audio-only visits are shown to be useful, an unexpected change to an audio-based visit from a planned video-based visit may result in slight inconvenience at best and a delay in essential visual assessment and therefore care at worst. In cases of communication breakdown, an in-person visit is often suggested.

With the patient's consent, providers should explore the option of having a patient's family member or friend assist with facilitating a virtual visit [13]. Patients and their families who are interested in virtual care should be provided with information and education about signing up for and using the service.

It is also helpful to provide some tips for the patient to prepare for the virtual care visit. The patient's clothing should allow a reasonable range of motion if the provider requests physical maneuvers during the virtual visit. If the patient requires a hearing or visual aid, these should be readily available. For patients with hearing, visual, or cognitive challenges, it is often more challenging to see and hear via a screen and speakers, in contrast to an in-person visit. Patients requiring mobility aids should keep these devices close by and may benefit from the presence of a nondisabled individual to assist them as necessary.

Etiquette

Many health care providers are not familiar with how their patients will perceive them in a virtual visit. With this in mind, it is useful to be aware of the etiquette that is unique to virtual care [42,43]. The basic principles of developing a high-quality "webside manner" can be separated into two areas: technical considerations and communication skills.

Technical Considerations

Due to technical limitations, auditory and visual information can be lost during a virtual consultation. At the beginning of the session, both cameras should be directed so that the faces and shoulders of both the patient and the health care provider are framed. Remember that slower, broader movements decrease visual blurring, and obscuring the face and mouth should be avoided. The physician should be mindful that gazing at the patient's image on the screen can make it seem as though the physician is looking down at them. When possible, the physician should look into the camera.

Communication Skills

During the session, the interviewer must recognize the importance of clear communication. Virtual visits require communication adjustments. Physicians should speak slowly and use nonmedical terms for maximum clarity. They should be prepared for lags in communication for technical reasons. Recognizing these technical limitations reinforces the need for speaking clearly. It may be helpful for the interviewer to check in with the patient to see if they can hear and understand their messaging. Reiterating and summarizing what a patient has said may also be helpful. In addition, clear signposting by the interviewer can indicate the progress and direction of the session. For example, the interviewer could say, "I am now going to summarize what I understand of your history, then proceed to the physical exam."

The primary sensory modalities available in a virtual visit are limited to auditory and visual. However, touch is a standard method of communication during an in-person visit, especially during the physical examination. When requesting that patients perform physical examination maneuvers during a virtual visit, adjustments are required. The clinician should clearly state the movement that the patient is asked to perform. They should consider documenting the absent physical examination maneuvers that would have typically been included but were not feasible due to the virtual nature of the consultation. This suggests, at a minimum, that these steps have been considered. Also, the clinician should inform the patient before they navigate to another screen to look at laboratory or imaging data. When possible, screen-sharing modes should be used to show patients their images or laboratory results.

Last but certainly not least, kindness and compassion are paramount in a virtual visit. Listening carefully to a patient's concerns, taking time to clarify their statements, and following up on their questions are vital skills, regardless of what form of medical care is being offered.

Environment and Setting

Preparing one's virtual visit environment is vital for both patients and providers. Clinical administrative staff play a key role. Prior to the appointment, the staff should send a message to the patient with recommendations on setting up for a visit. Some recommendations are relevant to patients and virtual visit providers alike:

- The room should be reasonably private, with minimal potential for distractions. It may be helpful to notify other members of the household or building that space will be used privately for the duration of the appointment. If possible, adequate lighting, presenting oneself in an uncluttered room, and ensuring the visibility of one's face and shoulders are ideal. Background should be reduced to a minimum. This may include distancing from a louder area (kitchen, public area, open windows, etc). We recommend posting a notice at the entrance to the room indicating that a confidential meeting is occurring. Personal communication devices (ie, mobile phones and pagers) should be switched to silent mode to avoid distraction.
- If other parties are involved in the meeting, such as trainees, family members, and friends, they should be introduced prior the beginning of the meeting. As with standard in-office visits, the people who are involved in the meeting should be transparent and within the control of the patient and physician.

Other recommendations are more relevant to patients and can be tailored to their medical conditions and clinical scenarios (eg, postoperative vs chronic disease follow-up):

- For some clinical specialties, it may be helpful for the patient's room to have enough space to allow them to stand and move away from the camera, so their entire body is visible.
- It is helpful if the camera is high-definition (720p or higher) and wide-angle. The room should have a comfortable yet easily movable chair available for the patient to rest.
- If the consultation requires equipment, the patient should have an easily accessible area to place these materials.
- A high-quality internet connection within the selected room is also essential. A wired connection is ideal, but if only Wi-Fi is available, the room should not be too far from the wireless router to ensure a reliable connection.

Example Run-through of a Virtual Visit

Presession Preparation

During the presession preparation, attention must be paid to the equipment, environment, and external appearance. In addition to checking the equipment and network, it should be ensured that the general guidelines laid out in the above section with regard to etiquette and the environment are followed. The patient's information or EHR should be available within reach of where the video visit is happening.

A good habit is to check the platform appointment software to confirm that the patient or patients have been booked as expected. If there is an issue with connecting to the patient, it is always useful to have an alternative means of communication. Ideally, this will be a telephone number; however, email or instant messaging can be used. Most platforms allow the patient to join and wait in a virtual “room.” It is recommended to mute the microphone before the patient enters the appointment.

Session

The first thing to do when the patient logs in is to confirm that both parties’ cameras and microphones are working correctly and troubleshoot as necessary. In all first visits, it should be expected that there will be delays due to technological issues. Technical factors play a significant role in one’s ability to communicate effectively with a patient. It is vital for the medical professional to be familiar with the platform they are using, as it is not uncommon to have to walk through teaching a patient how to modify camera or audio settings. It is essential to explain the steps in a simple and organized fashion.

Formal introductions can be made next, in which it is essential to obtain consent for the virtual visit and establish the identity of the patient. The ID verification step may not be necessary for clinical scenarios in which the patients are well known to the interviewer (eg, family practices, specialist practices managing chronic diseases). Announcing who is involved in the meeting for both parties is useful, as some participants can be virtual and others can be physical but not visible (off-camera).

The history will likely feel very similar to an in-person visit. It is useful to be clear and intentional with questions, as information can easily be lost through the virtual connection. The physical examination will be a different experience, as traditional medical training in conducting examinations anticipates that the patient will be physically present. When the correct group of patient, diagnosis, and visit purpose are selected, a virtual physical examination is equivalent to an in-person visit examination. Investigations can be reviewed with the patient, including sharing the actual results and images directly via screen sharing. If discussions with colleagues or other staff are necessary, the patient may be notified and muted. At the end of the appointment, as with most appointments, the clinician should discuss and establish a plan. Most relevant in this context is whether a follow-up appointment will be a virtual visit or in-person visit. Because the patient will not be taking any items (prescriptions, requisitions, instructions, etc) with them, it is crucial to verify details with the patient (eg, pharmacy, next steps).

Postsession

If orders and prescriptions are integrated into the EHR, clinicians can follow their routine workflow. Otherwise, orders and prescriptions must be sent post-encounter (eg, secure messaging, email, electronic prescriptions such as PrescribeIT, faxes, or mail). It should be ensured that follow-up arrangements have been made. Finally, the encounter can be documented by clinicians as per typical workflow (handwritten, dictation, or typed in the EHR), clearly identifying the visit as an eVisit or virtual visit.

Textbox 1 demonstrates a checklist that summarizes what we have discussed in the above sample run-through visit. It forms a guideline to maintain the quality of a virtual visit across all patients but also serves as a tool to structure future quality improvement projects.

Textbox 1. Checklist for virtual care.

Before the virtual video visit (presession):

- Check that the device, video conferencing equipment, and network are working properly
- Ensure that the camera and microphone are positioned properly
- Ensure that the room is well lit and does not contain any distractions and that only professional objects are visible
- Ensure privacy during a virtual video visit:
    - Mute your mobile phone or place it on vibrate
    - Close the office door and place a “Do Not Disturb” or “Doing Video Visits” sign
- Ensure beforehand that you have access to the patient charts, results, and any other health information
- Mute the microphone until the videoconference starts
- Ensure that the patient’s telephone number is available should it be necessary

During the virtual video visit:

- Confirm that both parties’ videoconferencing technology is working and optimized for sound and video
- Introduce yourself and bring your photo ID close to the camera for the patient to see
- Ask the patient to show a valid photo ID (driver’s license or Health Card)
- Obtain verbal consent for the virtual visit
- Ask the patient to ensure their room is well lit and has enough space for movement
- Ask the patient to ensure privacy during the virtual visit:
    - Mobile phone(s) are placed on mute or vibrate
    - Doors are closed and indicate “Do Not Disturb”
- Ask the patient to adjust their position so they are framed adequately within the camera view
- Ensure that the patient has all objects and tools necessary for the full assessment
- Introduce any other health care providers or learners present in the room

After finishing the virtual video visit:

- Ensure that all reports and clinical documentation are complete
- Ensure that the planned investigations and medication prescriptions are ordered
- Ensure that follow-up plans have been made

Conclusion

Virtual care will become a reality in many future medical practices, and the COVID-19 pandemic is serving as a catalyst for this impending change. Virtual care is effective at removing barriers to access, especially to specialists who are typically located in larger centers, for the remarkably distributed rural and remote patients across Canada. As we have outlined, there are important nuances in the application of virtual care that help ensure patients receive at least equivalent care, or in some ways, more effective care than that received in the traditional in-person format. Many clinical organizations are working toward establishing baseline principles, guidelines, and protocols to guide Canadian physicians. This paper is intended to contribute to an emerging Canadian body of knowledge and to support optimal professional practice in an emerging and rapidly evolving field. Most importantly, virtual care should not fragment care. It should support both high-quality care, the foundation of which is an established trust-based physician-patient relationship, and continuity of care.

Conflicts of Interest

None declared.

References

1. COVID-19 Pandemic Guidance for the Health Care Sector - Canada. Public Health Agency of Canada. 2020 Apr 22. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/covid-19-pandemic-guidance-health-care-sector.html#a24 [accessed 2020-12-17]
2. Bhyat R. Going Virtual in a Time of Crisis - Virtual Care resources for Canadian Physicians. LinkedIn. 2020 Mar 28. URL: https://www.linkedin.com/pulse/going-virtual-time-crisis-care-resources-canadian-physicians-bhyat [accessed 2020-12-17]
3. Virtual Care. OntarioMD. URL: https://ontariomd.news/ [accessed 2020-12-17]
4. Virtual Care: Expanded access to virtual care for all physicians. Ontario Medical Association. URL: https://www.oma.org/member/section/practice-&-professional-support/virtual-care?type=topics [accessed 2020-12-17]
5. Clifford T. Teladoc Health CEO: The health-care system is on the ‘verge of a new era’ for virtual care. CNBC. 2020 Mar 23. URL: https://www.cnbc.com/2020/03/23/teladoc-health-ceo-on-verge-of-a-new-era-for-virtual-health-care.html [accessed 2020-12-17]
6. Newsfile Corp. COVID-19: The Virtual Medical Doctor Will See You Now. Yahoo Finance. 2020 Mar 23. URL: https://finance.yahoo.com/news/covid-19-virtual-medical-doctor-144000742.html [accessed 2020-12-17]
7. CMA President says COVID-19 means we need to think not just about healthcare, but pandemic palliative care. CBC Sunday Magazine. 2020 Mar 20. URL: https://www.cbc.ca/radio/thesundayedition/the-sunday-edition-for-march-22-2020-1.5500395/cma-president-says-covid-19-means-we-need-to-think-not-just-about-healthcare-but-pandemic-palliative-care-1.5500776 [accessed 2020-12-17]
8. Shaw J, Jamieson T, Agarwal P, Griffin B, Wong I, Bhatia RS. Virtual care policy recommendations for patient-centred primary care: findings of a consensus policy dialogue using a nominal group technique. J Telemed Telecare 2018 Oct;24(9):608-615. [doi: 10.1177/1357633X17730444] [Medline: 28945161]
9. Virtual Care: Recommendations for Scaling Up Virtual Medical Services: Report of the Virtual Care Task Force. Canadian Medical Association. 2020 Feb. URL: https://www.cma.ca/sites/default/files/pdf/virtual-care/ReportoftheVirtualCareTaskForce.pdf [accessed 2020-12-17]
10. Lougheed T. Time to embrace the promise of virtual health care. CMAJ 2019 Mar 18;191(11):E320-E321 [FREE Full text] [doi: 10.1503/cmaj.109-5720] [Medline: 30885975]
11. Virtual Care in Canada: Discussion Paper. Canadian Medical Association. 2019. URL: https://www.cma.ca/sites/default/files/pdf/News/Virtual_Care_discussionpaper_v1EN.pdf [accessed 2020-12-17]
12. Hackett C, Brennan K, Smith Fowler H, Leaver C. Valuing Citizen Access to Digital Health Services: Applied Value-Based Outcomes in the Canadian Context and Tools for Modernizing Health Systems. J Med Internet Res 2019 Jun 06;21(6):e12277 [FREE Full text] [doi: 10.2196/12277] [Medline: 31172965]
13. Appireddy R, Khan S, Leaver C, Martin C, Jin A, Durafourt B, et al. Home Virtual Visits for Outpatient Follow-Up Stroke Care: Cross-Sectional Study. J Med Internet Res 2019 Oct 07;21(10):e13734 [FREE Full text] [doi: 10.2196/13734] [Medline: 31593536]
14. McGrail KM, Ahuja MA, Leaver CA. Virtual Visits and Patient-Centered Care: Results of a Patient Survey and Observational Study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
15. Paré G, Leaver C, Bourget C. Diffusion of the Digital Health Self-Tracking Movement in Canada: Results of a National Survey. J Med Internet Res 2018 May 02;20(5):e177 [FREE Full text] [doi: 10.2196/jmir.9388] [Medline: 29720359]
16. Enhanced Access to Primary Care: Project Evaluation Final Report. Women's College Hospital Institute for Health Systems Solutions and Virtual Care. 2020. URL: https://www-origin.otn.ca/wp-content/uploads/2019/08/eapc-evaluation-report.pdf [accessed 2020-12-17]
17. Khairat S, Meng C, Xu Y, Edson B, Gianforcaro R. Interpreting COVID-19 and Virtual Care Trends: Cohort Study. JMIR Public Health Surveill 2020 Apr 15;6(2):e18811 [FREE Full text] [doi: 10.2196/18811] [Medline: 32252023]
18. Jacobs J, Hu J, Slightam C, Gregory A, Zulman D. Virtual Savings: Patient-Reported Time and Money Savings from a VA National Telehealth Tablet Initiative. Telemed J E Health 2020 Sep;26(9):1178-1183. [doi: 10.1089/tmj.2019.0179] [Medline: 31880502]
19. Marcin J, Ellis J, Mawis R, Nagrampa E, Nesbitt T, Dimand R. Using telemedicine to provide pediatric subspecialty care to children with special health care needs in an underserved rural community. Pediatrics 2004 Jan;113(1 Pt 1):1-6. [doi: 10.1542/peds.113.1.1] [Medline: 14702439]
20. Coles N, Patel B, Li P, Cordeiro K, Steinberg A, Zdravkovic A, et al. Breaking barriers: Adjunctive use of the Ontario Telemedicine Network (OTN) to reach adolescents with obesity living in remote locations. J Telemed Telecare 2020 Jun;26(5):271-277. [doi: 10.1177/1357633X18816254] [Medline: 30526258]
21. Bryant T, Leaver C, Dunn J. Unmet healthcare need, gender, and health inequalities in Canada. Health Policy 2009 Jun;91(1):24-32. [doi: 10.1016/j.healthpol.2008.11.002] [Medline: 19070930]
22. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
23. Nouri S, Khoong E, Lyles C, Karliner L. Addressing Equity in Telemedicine for Chronic Disease Management During the Covid-19 Pandemic. NEJM Catalyst 2020 May 04 [FREE Full text] [doi: 10.1056/CAT.20.0123]

24. Crawford A, Serhal E. Digital Health Equity and COVID-19: The Innovation Curve Cannot Reinforce the Social Gradient of Health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
25. Khairat S, Haithcoat T, Liu S, Zaman T, Edson B, Gianforcaro R, et al. Advancing health equity and access using telemedicine: a geospatial assessment. J Am Med Inform Assoc 2019 Aug 01;26(8-9):796-805 [FREE Full text] [doi: 10.1093/jamia/ocz108] [Medline: 31340022]
26. Video visits enhance access to neurology follow up care. Ontario Telemedicine Network. 2019. URL: https://otn.ca/virtual-care-in-action/video-visits-for-stroke-follow-up/ [accessed 2020-12-17]
27. IIT Reacts Interactive audio-video platform. Reacts. URL: https://reacts.com/ [accessed 2020-12-17]
28. Bhyat R. Integrating digital health into medical education. Can Fam Physician 2019 Oct;65(10):683-686 [FREE Full text] [Medline: 31604728]
29. Telemedicine. College of Physicians and Surgeons of Ontario. 2014. URL: https://www.cpso.on.ca/Physicians/Policies-Guidance/Policies/Telemedicine [accessed 2020-12-17]
30. Covid-19: Virtual Care. College of Physicians and Surgeons of Alberta. URL: http://www.cpsa.ca/wp-content/uploads/2020/03/AP_COVID-19-Virtual-Care.pdf [accessed 2020-12-17]
31. Virtual Care Playbook for Canadian Physicians. Canadian Medical Association. 2020 Mar. URL: https://www.cma.ca/sites/default/files/pdf/Virtual-Care-Playbook_mar2020_E.pdf [accessed 2020-12-17]
32. Virtual Care Telephone and Video Codes Frequently Asked Questions. Ontario Medical Association. 2020 Mar 13. URL: https://www.hamiltonhealthsciences.ca/wp-content/uploads/2020/03/OMA-Telephone_and_Virtual_Care_Fee_Codes_FAQs_March13_2020.pdf [accessed 2020-12-17]
33. Adopting and Integrating Virtual Visits into Care: Draft Clinical Guidance For Health Care Providers in Ontario. Ontario Health. 2020 Mar 12. URL: https://quorum.hqontario.ca/Portals/0/Users/170/54/10154/Draft%20Clinical%20Guidance_Adopting%20and%20integrating%20virtual%20visits%20into%20care_V1.pdf [accessed 2020-12-17]
34. New guides for physicians on using virtual care. Doctors of BC. 2020 Mar 18. URL: https://www.doctorsofbc.ca/news/doctors-technology-office-virtual-care-support-response [accessed 2020-12-17]
35. Covid-19 Updates 2020 updated March 15, 2020. Doctors Manitoba. 2020 Mar 15. URL: https://doctorsmanitoba.ca/2020/03/covid-19-coronavirus-updates/ [accessed 2020-12-17]
36. Telemedicine and virtual care guidelines (and other clinical resources for COVID-19). Royal College of Physicians and Surgeons of Canada. 2020. URL: http://www.royalcollege.ca/rcsite/documents/about/covid-19-resources-telemedicine-virtual-care-e [accessed 2020-12-17]
37. COVID-19 Hub: Advice, Support and Medical-legal information. Canadian Medical Protective Association. 2020 Apr 03. URL: https://www.cmpa-acpm.ca/en/covid19 [accessed 2020-12-17]
38. Thinking of working with virtual clinics? Consider these medical-legal issues. Canadian Medical Protective Association. 2018 Jul. URL: https://www.cmpa-acpm.ca/en/advice-publications/browse-articles/2018/thinking-of-working-with-virtual-clinics---consider-these-medical-legal-issues [accessed 2020-12-17]
39. Ruotsalainen P, Blobel B. Digital pHealth - Problems and Solutions for Ethics, Trust and Privacy. Stud Health Technol Inform 2019;261:31-46. [Medline: 31156088]
40. Advanced and Meaningful Use of EMRs. College of Family Physicians of Canada and Canada Health Infoway. 2018. URL: https://www.infoway-inforoute.ca/en/component/edocman/3591-best-advice-advanced-and-meaningful-use-of-emrs/view-document?Itemid=0 [accessed 2020-12-17]
41. Electronic Records Handbook. Canadian Medical Protective Association. 2020. URL: https://www.cmpa-acpm.ca/static-assets/pdf/advice-and-publications/handbooks/com_electronic_records_handbook-e.pdf [accessed 2020-12-17]
42. Iyengar M, Rogith D, Florez-Arango J. Measuring Workload Demand of Informatics Systems with the Clinical Case Demand Index. AMIA Annu Symp Proc 2017;2017:985-993 [FREE Full text] [Medline: 29854166]
43. How to set up virtual care in your practice. Canadian Medical Association. How to set up virtual care in your practice: Canadian Medical Association URL: https://www.cma.ca/how-set-virtual-care-your-practice [accessed 2020-07-10]

Abbreviations

CMPA: Canadian Medical Protective Association
EHR: electronic health record
PIPEDA: Personal Information Protection and Electronics Document Act
PHIPA: Personal Health Information Protection Act

Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by S Bhatia, N Khalili-Mahani, O Abiodun-Ojo; comments to author 22.06.20; revised version received 29.07.20; accepted 02.10.20; published 21.01.21.

Please cite as:
Wong A, Bhyat R, Srivastava S, Boissé Lomax L, Appireddy R
Patient Care During the COVID-19 Pandemic: Use of Virtual Care
J Med Internet Res 2021;23(1):e20621
URL: https://www.jmir.org/2021/1/e20621
doi:10.2196/20621
PMID:33326410

Original Paper

Social, Cognitive, and eHealth Mechanisms of COVID-19–Related Lockdown and Mandatory Quarantine That Potentially Affect the Mental Health of Pregnant Women in China: Cross-Sectional Survey Study

Xue Yang[1], PhD; Bo Song[2], MMed; Anise Wu[3], PhD; Phoenix K H Mo[1], PhD; Jiangli Di[2], PhD; Qian Wang[2], PhD; Joseph T F Lau[1], PhD; Linhong Wang[4], MMed

[1]The Jockey Club School of Public Health and Primary, The Chinese University of Hong Kong, Hong Kong, China

[2]National Center for Women and Children's Health, Chinese Center for Disease Control and Prevention, Beijing, China

[3]Department of Psychology, University of Macau, Macao, Macao

[4]National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

Corresponding Author:
Qian Wang, PhD
National Center for Women and Children's Health
Chinese Center for Disease Control and Prevention
No 12 Dahuisi Road
Haidan District
Beijing, 100081
China
Phone: 86 01062170871
Email: qianawang@chinawch.org.cn

Abstract

Background: Although lockdown and mandatory quarantine measures have played crucial roles in the sharp decrease of the number of newly confirmed/suspected COVID-19 cases, concerns have been raised over the threat that these measures pose to mental health, especially the mental health of vulnerable groups, including pregnant women. Few empirical studies have assessed whether and how these control measures may affect mental health, and no study has investigated the prevalence and impacts of the use of eHealth resources among pregnant women during the COVID-19 outbreak.

Objective: This study investigated (1) the effects of lockdown and mandatory quarantine on mental health problems (ie, anxiety and depressive symptoms), (2) the potential mediation effects of perceived social support and maladaptive cognition, and (3) the moderation effects of eHealth-related factors (ie, using social media to obtain health information and using prenatal care services during the COVID-19 pandemic) on pregnant women in China.

Methods: An online cross-sectional survey was conducted among 19,515 pregnant women from all 34 Chinese provincial-level administrative regions from February 25 to March 10, 2020.

Results: Of the 19,515 participants, 12,209 (62.6%) were subjected to lockdown in their areas of residence, 737 (3.8%) were subjected to mandatory quarantine, 8712 (44.6%) had probable mild to severe depression, 5696 (29.2%) had probable mild to severe anxiety, and 1442 (7.4%) had suicidal ideations. Only 640 (3.3%) participants reported that they used online prenatal care services during the outbreak. Significant sociodemographic/maternal factors of anxiety/depressive symptoms included age, education, occupation, the area of residence, gestational duration, the number of children born, complication during pregnancy, the means of using prenatal care services, and social media use for obtaining health information. Multiple indicators multiple causes modeling (χ^2_{14}=495.21; *P*<.05; comparative fit index=.99; nonnormed fit index=.98; root mean square error of approximation=.04, 90% CI 0.038-0.045) showed that quarantine was directly and indirectly strongly associated with poor mental health through decreased perceived social support and increased maladaptive cognition (B=.04; β=.02, 95% CI 0.01-0.02; *P*=.001), while lockdown was indirectly associated with mental health through increased social support and maladaptive cognition among pregnant women (B=.03; β=.03, 95% CI 0.02-0.03; *P*=.001). Multigroup analyses revealed that the use of social media for obtaining health information and the means of using prenatal care services were significant moderators of the model paths.

Conclusions: Our findings provide epidemiological evidence for the importance of integrating mental health care and eHealth into the planning and implementation of control measure policies. The observed social and cognitive mechanisms and moderators in this study are modifiable, and they can inform the design of evidence-based mental health promotion among pregnant women.

(J Med Internet Res 2021;23(1):e24495) doi:10.2196/24495

KEYWORDS

eHealth; lockdown; quarantine; depression; anxiety; pregnant women

Introduction

Background

Lockdown and mandatory quarantine are commonly used and effective measures that are implemented by governments to contain the transmission of respiratory infectious diseases, including the COVID-19 disease [1,2]. Lockdown refers to general and widespread restrictions on movement, work, and travel for all people in a city, region, or country. Lockdown measures include travel restrictions, the mandatory closure of schools, and bans on nonessential commercial and social activities. Given the global spread of COVID-19 to 216 countries/regions, which has resulted in over 30.6 million infections and 950,000 deaths as of September 20, 2020 [3], more than 100 countries (eg, the United States, France, Australia, Thailand, and South Africa) have adopted various forms of lockdown measures to control the pandemic [4]. In mainland China, more than 80 cities in around 20 provinces and municipalities were put in lockdown [5], and over 780 million people were under certain travel restrictions [6].

Mandatory quarantine is a form of isolating people who are not ill, but may have been exposed to a disease that is dangerous to society [7]. People who had close contact with individuals who were confirmed/suspected to have COVID-19, and people with a history of travel were quarantined for 14 days in designated facilities (eg, governmental facilities and hotels) or home settings [8]. Globally, more than 140 countries (eg, the United States, the United Kingdom, Spain, Italy, and Singapore) have adopted mandatory quarantine measures for disease control [9-12].

The Impact of Lockdown and Mandatory Quarantine on Mental Health

Although lockdown and mandatory quarantine have played crucial roles in the sharp decrease of the number of newly confirmed/suspected COVID-19 cases [13-15], concerns have been raised over the threat that these measures pose to mental health, as these unprecedented measures have restricted daily routines and increased social isolation [16,17]. Empirical studies on the impact of lockdown/mandatory quarantine on mental health have yielded inconsistent results. According to the Office of National Statistics, more than 25 million people in the United Kingdom have experienced high levels of anxiety in late March 2020, which is when the lockdown was announced [17]. In a study conducted by Sibley et al [18], participants from New Zealand reported a slight increase in psychological distress, but less fatigue and no significant changes in rumination, feelings of belongingness, perceived social support, satisfaction with life, standards of living, future security, personal relationships, and health during the early phase of the nationwide lockdown compared to those during the pre-COVID-19 period. A study that was conducted in southern China during early February 2020 also found that people who were quarantined had a greater prevalence of anxiety and depression than those who were not affected by quarantine [19]. Furthermore, in mid-February 2020, Li et al [20] reported a positive association between perceived inconvenience to daily life caused by home quarantine and depression/anxiety among the general adult population in China. However, Zhu et al [21] reported that there was no significant differences in depression and anxiety between Chinese people who were and were not subjected to mandatory quarantine, and they concluded that although these mental health problems were not related to quarantine control measures, these measures did impact daily life. Another study in China even found a significantly lower prevalence of depression and anxiety among people under mandatory quarantine than among those who were infected by SARS-CoV-2 or the general public [22]. These inconsistent results highlight the importance of exploring potential underlying mechanisms (eg, the impacts of lockdown and quarantine on individuals) that may explain the relationship between lockdown and quarantine measures and mental health problems. We however did not identify such studies.

Pregnant women could be more susceptible to lockdown and quarantine measures and SARS-CoV-2 infection than the general population, due to their great need for social support and ongoing prenatal care services, concerns over fetal safety, immunocompromised status, and physiological and psychosocial adaptive changes during pregnancy [23]. Most studies on the mental health of pregnant women have small sample sizes (ie, 70 participants to around 560 participants), and these studies have only reported the prevalence of mental health problems during the COVID-19 pandemic [24-37]. We found 3 studies on the mental health of pregnant women that had larger sample sizes, with about ≥1000 respondents (ie, 946 participants to around 2421 participants) [24,25,30], and only 1 study (N=260) that investigated the impact of social isolation on mental health by simply asking pregnant women whether they believed that social isolation due to the pandemic affected their psychological well-being [27]. More studies that include large sample sizes and investigate the impact of lockdown/mandatory quarantine on the mental health of pregnant women during the COVID-19 pandemic are needed. Early detection and intervention can prevent the adverse impact that mental problems (eg, prenatal mental disorders) have on both mothers and children in the long term.

Potential Social and Cognitive Mechanisms

Social and cognitive mechanisms may play critical roles in mediating the relationships between lockdown/mandatory quarantine and mental health. First, lockdown and quarantine are isolation measures that, by their nature, may induce social isolation and reduce social resources, such as social support, and reductions in such social resources are a risk factor of mental health problems [16]. We found 2 studies with a sample size that ranged between 308 participants to around 403 participants. These studies investigated the association between social support and mental health problems among pregnant women during the COVID-19 pandemic, and both studies reported negative associations [26,28]. The mediation effects of social support are supported by the conservation of resources theory, which predicts that resource loss (eg, losses of social resources like social support) is the principal factor in the stress process and the cause of mental disorder development [38].

Second, the governmental implementation of unprecedented measures for disease control may be a stressful event for the public that induces maladaptive cognitive responses. According to the response styles theory, both depression and anxiety are related to faulty cognitive responses to stressors and negative emotions [39,40]. Rumination (ie, the repetition of the same feelings and thoughts) and catastrophizing (ie, having thoughts that explicitly emphasize the terror of what one has experienced) are common maladaptive cognitions, and responses to maladaptive cognitions emerge when individuals experience threatening and uncertain events [41]. When compared to the different kinds of maladaptive cognitive responses, anxiety has been found to be more related to catastrophizing, which focuses on future threats [42], and depression has been found to be more related to ruminative thinking, which concentrates on past negative experiences and emotions [43,44]. Empirical studies have also supported the mediation roles of rumination and catastrophizing between threatening events (eg, daily hassles) and depression/anxiety [41,45]. No study has assessed the mediation effects of rumination and catastrophizing in the context of COVID-19 or pregnant women.

eHealth-Related Moderators

eHealth refers to information and communications technologies in health care and the community. eHealth can be an optimal communication modality for people under stay-at-home orders, especially for those with time-sensitive health conditions, such as pregnancy [46]. eHealth and telemedicine services enable pregnant women to maintain their regular prenatal visit schedule and avoid the unnecessary risk of COVID-19 exposure [47]. Such online services may significantly affect individuals' coping resources, stress appraisals, and perceived or actual social support [48,49], and may reduce the influence of external and environmental stress on individuals' mental health during the COVID-19 pandemic. Thus, we hypothesized that the effects of control measures on interpersonal resources (eg, social support), cognitive status (eg, maladaptive cognition), and mental health might vary between pregnant women who use eHealth resources, such as using social media to obtain health information and online services to make appointments with doctors for prenatal care services during the COVID-19 pandemic, and those who did not use such resources. In addition, such extra resources may reduce the adverse effects of maladaptive cognition and enhance the protective effects of social support on mental health. We did not find any research that investigated the prevalence of eHealth resource use among pregnant women and the impact of using eHealth resources on the mental health of pregnant women.

Objectives

This study aimed to investigate the prevalence of mental health problems (ie, depression and anxiety) among a large sample of pregnant women recruited from multiple regions in mainland China. We also assessed the direct and indirect effects of lockdown and mandatory quarantine on mental health problems through perceived social support and maladaptive cognition. We hypothesized that lockdown (ie, model path H1) and mandatory quarantine (ie, model path H2) would be positively associated with mental health problems. We further hypothesized that lockdown and mandatory quarantine would be indirectly associated with mental health problems through reduced perceived social support (ie, model path H3) and increased maladaptive cognition (ie, model path H4). Moreover, this study aimed to assess the moderation effects of eHealth-related variables, including using social media to obtain health information, using online prenatal care services, and making appointments with doctors during the COVID-19 outbreak, on pregnant women for each model path.

Methods

Participant Recruitment and Procedure

In this study, the inclusion criteria for the sample were (1) female sex, (2) age ≥18 years, (3) the ability to speak Chinese, (4) current pregnancy, and (5) the use of maternal health care services provided by the Maternal and Child Health Hospitals of the Chinese Preventive Medicine Association. Pregnant women who planned to terminate their pregnancy were excluded from this study. An online cross-sectional survey was conducted from February 24 to March 10, 2020. Eligible participants were identified from the records of Maternal and Child Health Hospitals from multiple regions of China and were invited to take part in the online survey by doctors through WeChat (Tencent Inc), which is the most widely used social media platform in Chinese populations. Interested participants visited the online survey through a link or quick response code and read the informed consent form before starting the survey. They were informed that clicking the "starting the survey" button implied informed consent. They were also informed that the study was anonymous and confidential, and that refusal to take part in the survey would not affect any services they would obtain. The survey took about 15 minutes to complete. No incentive was provided. A total of 19,515 pregnant women from all 34 provincial-level administrative regions in China (eg, 2127 pregnant women from Beijing, 4015 from Shandong, 3659 from Zhejiang, 1886 from Guangdong, 1250 from Hunan, 3178 from Shanxi, etc) completed the survey, with a valid response rate of 87.7%. This study was approved by the Survey and Behavioural Research Ethics Committee of the Chinese University of Hong Kong (Number SBRE-19-395).

Measures

Sociodemographic and maternal information, including age, education level, occupation, the area of residence, gestational duration, the number of children born, complication during pregnancy, the major means of using prenatal care service during the COVID-19 pandemic (ie, using online services, making appointments with doctors, and going to a hospital as usual), and the frequency of using social media to obtain health information in the past week, were obtained from the survey. Participants also reported on their lockdown and mandatory quarantine status by answering the following questions: (1) “Had the city, town, or county where you currently live been put under lockdown by the local government because of the COVID-19 epidemic” (response score: 0=no and 1=yes); and (2) “Had you been under mandatory quarantine (e.g., governmental facilities-, hotel- or home-quarantine) because of the COVID-19 epidemic” (response score: 0=no and 1=yes)?

Perceived social support was measured by 2 dimensions (ie, general social support and perceived change in social support during the COVID-19 pandemic compared to those during the pre-COVID-19 period). The following 2 survey items were used to assess this: (1) “Overall, to what extent did you receive social support from families, friends, and others during the COVID-19 epidemic” (response scale: 1=very poor to 10=very good); and (2) “To what extent did your social support become poorer or better during COVID-19, compared to that before the outbreak of COVID-19” (response scale: 1=much poorer to 5=much better)? Similar survey items have been used in previous studies [50].

Maladaptive cognition related to COVID-19 was measured with the short-form Cognitive Emotion Regulation Questionnaire (CERQ) [51]. The CERQ was developed to evaluate the cognitive aspects of emotion regulation when one experiences stressful or unpleasant events. Sample items from the CERQ include “I am preoccupied with what I think and feel about what I have experienced” and “I keep thinking about how terrible it is what I have experienced.” These items are rated on Likert scales (ie, 1=almost never to 5=almost always). The subscales of rumination and catastrophizing were used in this study. The Chinese version has been validated in previous studies [52]. The reliability of the 2 subscales was acceptable for our sample (Cronbach α=.66; Cronbach α=.84, respectively).

Depression was measured using the Patient Health Questionnaire-9 (PHQ-9) [53]. Respondents evaluated the presence (PHQ-9 response scale: 0=none to 3=almost every day) of 9 criteria for a depressive episode that occurred in the past 2 weeks, in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (eg, “Feeling bad about yourself—or that you are a failure or have let yourself or your family down?”). The cutoff points for mild, moderate, moderately severe, and severe depression were total PHQ-9 scores of 5, 10, 15, and 20, respectively. The Chinese version has been used in previous studies [54], and it had good internal consistency (Cronbach α=.83).

Anxiety was measured by the Generalized Anxiety Disorder 7-item (GAD-7) scale [55]. It is based on DSM-IV criteria and is used to measure the severity of generalized anxiety disorder based on the past 2 weeks. Participants respond according to a 4-point Likert-type scale (ie, 0=none to 3=almost every day). The cutoff points for mild, moderate, and severe anxiety were total GAD-7 scores of 5, 10, and 15 respectively. The Chinese version has been validated in previous studies [56]. It had a Cronbach α of .90 with our sample.

Data Analyses

Descriptive statistics, including frequency, means, and standard deviations, were computed for participants’ sociodemographic characteristics. Differences in depression scores and anxiety scores based on sociodemographic characteristics were compared by either an independent 2-tailed *t* test or analysis of variance. The effect size (ie, Cohen *d* or Cohen *f*) was reported. Bivariate correlations between the key variables were presented. The effect size was considered low if the value of *r* varied around .10, medium if *r* varied around .30, and large if *r* varied by more than .50 [57]. Multiple indicators multiple causes (MIMIC) modeling was conducted to test the proposed mediation model. The goodness of fit was tested, and standardized path coefficients (ie, β) were reported. The mediation hypotheses were tested by bootstrapping analyses. The 95% confidence intervals of the indirect effects were obtained from 5000 bootstrap samples. The effect size (ie, proportion of mediation [PM]) was reported. Multigroup analyses were conducted to test the proposed moderators; *P*<.05 in the Chi-square difference test ($\Delta\chi^2/\Delta df$) would suggest a significant moderation effect. The missing data rate was below 5%, and all missing values were replaced by using multiple imputation. The level of statistical significance was .05, and SPSS version 21.0 (IBM Corp) and Amos Version 26 (IBM Corp) were used for data analyses.

Results

Sociodemographic and Maternal Characteristics

Tables 1-3 present the sociodemographic and maternal characteristics of the participants. Of 19,515 participants, 13,885 (71.1%) were aged 26-35 years, 11,627 (59.6%) had an education level of college or above, 10,741 (55%) did not have a baby before this pregnancy, and 17,856 (91.5%) did not experience any complications or comorbidities during pregnancy. Additionally, 5394 (27.6%) participants were unemployed/housewives (27.6%) and 5193 (26.6) were technical/administrative personnel (26.6%). The distribution of participants’ areas of residence was approximately even (municipality/provincial capitals: n=5930, 30.4%; general cities: n=6855, 35.1%; counties: n=6730, 34.5%). Participants’ average gestational duration was 25.4 weeks (SD 9.8 weeks).

Table 1. Sociodemographic and maternal characteristics of the participants (N=19,515).

Characteristics	Value
Age (years), n (%)	
<26	3781 (19.4)
26-30	8202 (42)
31-35	5683 (29.1)
>35	1849 (9.5)
Education, n (%)	
Middle school or below	4014 (20.6)
High school	3874 (19.9)
College	5222 (26.8)
University or above	6405 (32.8)
Occupation, n (%)	
Technical personnel	3053 (15.6)
Administrative personnel	2140 (11)
Civil servant	320 (1.6)
Soldier	22 (0.1)
Business/service personnel	1778 (9.1)
Self-employed/private business owner	1455 (7.5)
Farmer/migrant worker	746 (3.8)
Unemployed/housewife	5394 (27.6)
Student (undergraduate/postgraduate)	51 (0.3)
Other	4556 (23.3)
Area of residence, n (%)	
Municipality/provincial capital	5930 (30.4)
General city	6855 (35.1)
County	6730 (34.5)
Number of gestational weeks, mean (SD)	25.4 (9.8)
Gestational duration (weeks), n (%)	
1-10	1523 (7.8)
11-20	4986 (25.5)
21-30	5858 (30)
>30	6518 (33.4)
Not sure	630 (3.2)
Number of children born before pregnancy, n (%)	
0	10741 (55)
1	7796 (39.9)
>1	978 (5)
Complication during pregnancy, n (%)	
Yes	1659 (8.5)
No	17856 (91.5)

Table 2. Sociodemographic and maternal characteristics of the participants (N=19,515) based on eHealth-related variables.

eHealth-related variables	Value
Means of using prenatal care service, n (%)	
Went to hospital as usual	10189 (52.2)
Made appointment with doctor	7568 (38.8)
Online prenatal care	640 (3.3)
Not sure	1118 (5.7)
Used social media to obtain health information, n (%)	
Never	1781 (9.1)
Sometimes	10605 (54.3)
Always	7129 (36.6)

With regard to the means of using prenatal care services during the COVID-19 epidemic, 10,189 (52.2%) participants went to the hospital for prenatal care as usual, 7568 (38.8%) made appointments with doctors, 640 (3.3%) used online services, and 1118 (5.7%) were uncertain. Most participants (n=17,734, 90.9%) used social media to obtain health information in the week before the survey.

Of the 19,515 participants, 12,209 (62.6%) participants reported lockdown in their areas of residence; 737 (3.8%) were subjected to mandatory quarantine; 8712 (44.6%) had probable mild to severe depression; 1442 (7.4%) had suicidal ideations, as measured by question 9 in the PHQ-9; and 5696 (29.2%) had probable mild to severe anxiety.

Table 3. Sociodemographic and maternal characteristics of the participants (N=19,515) based on psychosocial variables.

Psychosocial variables	Value
Lockdown in the area of residence, n (%)	
Yes	12209 (62.6)
No	7306 (37.4)
Subjected to quarantine, n (%)	
Yes	737 (3.8)
No	18778 (96.2)
General social support, mean (SD)	8.51 (2.07)
Social support change, mean (SD)	4.05 (1.06)
Rumination, mean (SD)	3.11 (0.86)
Catastrophizing, mean (SD)	2.76 (1.09)
Depressive symptoms, mean (SD)	0.56 (0.56)
Depressive symptoms (total PHQ-9[a] score), n (%)	
Minimal (0-4)	10803 (55.4)
Mild (5-9)	5565 (28.5)
Moderate (10-14)	2053 (10.5)
Moderately severe (15-19)	793 (4.1)
Severe (20-27)	301 (1.5)
Anxiety symptoms, mean (SD)	0.46 (0.60)
Anxiety symptoms (total GAD-7[b] score), n (%)	
Minimal (0-4)	13819 (70.8)
Mild (5-9)	4177 (21.4)
Moderate (10-14)	1052 (5.4)
Severe (15-21)	467 (2.4)
Self-harm/suicidal ideation[c], mean (SD)	0.10 (0.41)
Self-harm/suicidal ideation frequency[c], n (%)	
None	18073 (92.6)
Several days	979 (5.0)
More than half of the days	325 (1.7)
Almost every day	138 (0.7)

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

[b] GAD-7: Generalized Anxiety Disorder 7-item.

[c] Based on item 9 in the Patient Health Questionnaire-9.

Level of Mental Health Problems Based on Sociodemographic Characteristics

As seen in Tables 4 and 5, factors that positively significantly associated with both depressive and anxiety symptoms included young age, being a student, residing in counties, being in the early or final stages of pregnancy, using means of prenatal care other than making appointments with doctors, using social media to obtain health information, experiencing lockdown in the areas of residence, and being quarantined. Other factors that were significantly associated with greater anxiety included low education levels (*P*=.01) and complications during pregnancy (*P*<.001). With regard to the number of children born, participants who were giving birth for the first time reported significantly more depressive symptoms than those who were not giving birth for the first time (*P*<.001). They also reported more anxiety symptoms than those who had given birth to 1 child before this pregnancy (*P*=.002).

Table 4. Depressive and anxiety symptoms stratified by sociodemographic and maternal characteristics.

Variables	Depressive symptoms				Anxiety symptoms			
	Mean (SD)	*F* test (df) or *t* test (df)	*P* value	Cohen *d* or Cohen *f*	Mean (SD)	*F* test (df) or *t* test (df)	*P* value	Cohen *d* or Cohen *f*
Age (years)	N/A[a]	34.95 (319,511)	<.001	0.07	N/A	18.12 (319,511)	<.001	0.05
<26	0.62 (0.62)				0.51 (0.67)			
26-30	0.56 (0.54)				0.45 (0.58)			
31-35	0.54 (0.54)				0.45 (0.59)			
>35	0.57 (0.52)				0.40 (0.57)			
Education	N/A	2.30 (319,511)	.08	0.02	N/A	4.17 (319,511)	.01	0.03
Middle school or below	0.55 (0.61)				0.48 (0.66)			
High school	0.58 (0.60)				0.47 (0.63)			
College	0.56 (0.54)				0.44 (0.57)			
University or above	0.55 (0.51)				0.45 (0.57)			
Occupation	N/A	3.80 (919,505)	<.001	0.04	N/A	3.35 (919,505)	<.001	0.04
Technical staff	0.55 (0.51)				0.44 (0.57)			
Administrative staff	0.54 (0.51)				0.44 (0.57)			
Civil servant	0.62 (0.56)				0.52 (0.66)			
Soldier	0.45 (0.45)				0.32 (0.40)			
Business/service personnel	0.57 (0.55)				0.45 (0.58)			
Self-employed/private business owner	0.56 (0.57)				0.45 (0.59)			
Farmer/migrant worker	0.51 (0.60)				0.43 (0.63)			
Unemployed/housewife	0.58 (0.59)				0.48 (0.63)			
Student (undergraduate/postgraduate)	0.75 (0.66)				0.69 (0.71)			
Other	0.54 (0.56)				0.45 (0.60)			
Area of residence	N/A	12.05 (219,512)	<.001	0.04	N/A	8.24 (219,512)	<.001	0.03
Municipality/provincial capital	0.53 (0.52)				0.44 (0.47)			
General city	0.44 (0.56)				0.45 (0.60)			
County	0.58 (0.59)				0.48 (0.63)			
Gestational duration (week)	N/A	10.64 (419,510)	<.001	0.04	N/A	9.71 (419,510)	<.001	0.04
1-10	0.62 (0.61)				0.48 (0.64)			
11-20	0.57 (0.55)				0.44 (0.59)			
21-30	0.53 (0.54)				0.43 (0.58)			
>30	0.56 (0.55)				0.48 (0.61)			
Not sure	0.56 (0.67)				0.54 (0.73)			
Number of children born before pregnancy	N/A	17.01 (219,512)	<.001	0.04	N/A	6.34 (219,512)	.002	0.02
0	0.58 (0.54)				0.47 (0.60)			
1	0.53 (0.57)				0.44 (0.60)			
>1	0.52 (0.63)				0.48 (0.80)			
Complication during pregnancy	N/A	0.03 (19,513)	0.98	0.00	N/A	3.70 (1961)	<.001	0.10
Yes	0.56 (0.55)				0.51 (0.62)			

Variables	Depressive symptoms				Anxiety symptoms			
	Mean (SD)	*F* test (df) or *t* test (df)	*P* value	Cohen *d* or Cohen *f*	Mean (SD)	*F* test (df) or *t* test (df)	*P* value	Cohen *d* or Cohen *f*
No	0.56 (0.56)				0.45 (0.60)			

[a] N/A: not applicable.

Table 5. Depressive and anxiety symptoms stratified by eHealth-related variables.

Variables	Depressive symptoms				Anxiety symptoms			
	Mean (SD)	*F* test (df) or *t* test (df)	*P* value	Cohen *d* or Cohen *f*	Mean (SD)	*F* test (df) or *t* test (df)	*P* value	Cohen *d* or Cohen *f*
Means of using prenatal care services	N/A[a]	29.35 (319,511)	<.001	0.06	N/A	23.66 (319,511)	<.001	0.06
Went to hospital as usual	0.57 (0.56)				0.46 (0.61)			
Made appointment with doctor	0.52 (0.53)				0.42 (0.57)			
Online prenatal care	0.60 (0.62)				0.50 (0.65)			
Not sure	0.67 (0.63)				0.58 (0.69)			
Used social media to obtain health information	N/A	3.69 (219,512)	.03	0.03	N/A	3.06 (219,512)	.04	0.03
Never	0.52 (0.61)				0.43 (0.66)			
Sometimes	0.56 (0.55)				0.45 (0.59)			
Always	0.56 (0.55)				0.47 (0.60)			

[a] N/A: not applicable.

Bivariate Correlations Between the Key Psychosocial Variables

As seen in Table 6, living in an area under lockdown had significant positive associations with all the key psychological variables, including perceived general social support (*P*<.001), social support change (*P*<.001), rumination (*P*<.001), catastrophizing (*P*<.001), depression (*P*=.01), and anxiety (*P*=.03). Being quarantined was significantly and negatively associated with perceived general social support (*P*<.001) and social support change (*P*=.01), while it was significantly and positively associated with catastrophizing (*P*<.001), depression (*P*<.001), and anxiety (*P*<.001). However, all the correlations associated with lockdown and quarantine had small effect sizes. Both perceived general social support and social support change had small to moderate negative correlations with depression (*r*=−0.17, *P*<.001; *r*=−0.16, *P*<.001, respectively) and anxiety (*r*=−0.17, *P*<.001; *r*=−0.15, *P*<.001, respectively). Both rumination and catastrophizing had moderate positive correlations with depression (*r*=0.25, *P*<.001; *r*=0.28, *P*<.001, respectively) and anxiety (*r*=0.26, *P*<.001; *r*=0.31, *P*<.001, respectively).

Table 6. Bivariate correlations (Spearman ρ, Pearson *r*, and *P* values) between the key psychosocial variables.

Variable	Lockdown	Quarantine	General social support	Social support change	Rumination	Catastrophizing	Depressive symptoms	Anxiety symptoms
Lockdown								
ρ	1	0.06[a]	0.04[a]	0.06[a]	0.05[a]	0.09[a]	0.02[b]	0.02[c]
P value	—[d]	<.001	<.001	<.001	<.001	<.001	.01	.03
Quarantine								
ρ	0.06[a]	1	−0.03[a]	−0.02[b]	0.01	0.03[a]	0.04[a]	0.04[a]
P value	<.001	—	<.001	.01	.12	<.001	<.001	<.001
General social support								
r	—	—	1	0.64[a]	0.05[a]	−0.04[a]	−0.17[a]	−0.17[a]
P value	—	—	—	<.001	<.001	<.001	<.001	<.001
Social support change								
r	—	—	0.64[a]	1	0.09[a]	0.02[c]	−0.16[a]	−0.15[a]
P value	—	—	<.001	—	<.001	.02	<.001	<.001
Rumination								
r	—	—	0.05[a]	0.09[a]	1	0.66[a]	0.25[a]	0.26[a]
P value	—	—	<.001	<.001	—	<.001	<.001	<.001
Catastrophizing								
r	—	—	−0.04[a]	0.02[c]	0.66[a]	1	0.28[a]	0.31[a]
P value	—	—	<.001	.02	<.001	—	<.001	<.001
Depressive symptoms								
r	—	—	−0.17[a]	−0.16[a]	0.25[a]	0.28[a]	1	0.78[a]
P value	—	—	<.001	<.001	<.001	<.001	—	<.001
Anxiety symptoms								
r	—	—	−0.17[a]	−0.15[a]	0.26[a]	0.31[a]	0.78[a]	1
P value	—	—	<.001	<.001	<.001	<.001	<.001	—

[a] The correlation is significant at a level of .001.

[b] The correlation is significant at a level of .01.

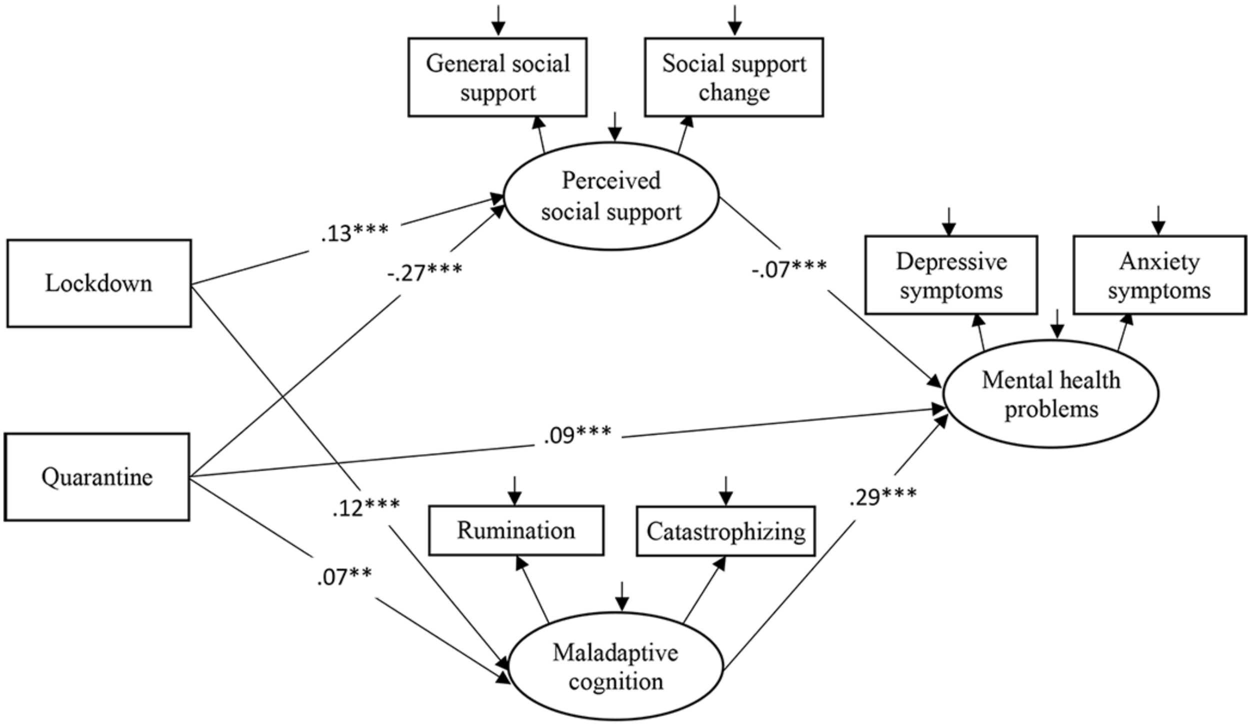
[c] The correlation is significant at a level of .05.

[d] Not applicable.

MIMIC Modeling Analyses for the Proposed Model

MIMIC modeling showed that both the measurement model (χ^2_{12}=411.75; comparative fit index [CFI]=.99; nonnormed fit index [NNFI]=.98; root mean square of error approximation [RMSEA]=.04, 90% CI 0.038-0.045) and structural model (χ^2_{14}=495.21; CFI=.99; NNFI=.98; RMSEA=.04, 90% CI 0.038-0.045) (Figure 1) fit the data well. Lockdown was significantly and positively associated with perceived social support (B=.13; β=.04; *P*<.001) and maladaptive cognition (B=.12; β=.09; *P*<.001), but not with mental health problems (B=.01; β=.01; *P*=.33). Mandatory quarantine was associated with low perceived social support (B=−.27; β=−.03; *P*<.001), great maladaptive cognition (B=.07, β=.02, *P*=.01), and great mental health problems (B=.09; β=.04; *P*<.001). Perceived social support was negatively associated with mental health problems (B=−.07; β=−.23; *P*<.001), while maladaptive cognition was positively associated with mental health problems (B=.29; β=.39; *P*<.001). Detailed results regarding regression weights are provided in Multimedia Appendix 1.

Figure 1. Proposed mediation model with unstandardized path coefficients. The nonsignificant path is not shown for simplicity reasons. **P*<.05, ***P*<.01, ****P*<.001.



Mediation Test

The indirect effects of lockdown (B=.03; β=.03, 95% CI 0.02-0.03; *P*=.001; PM=78.8%) and quarantine (B=.04; β=.02, 95% CI 0.01-0.02; *P*=.001; PM=30.1%) on mental health problems were statistically significant. Specifically, the indirect effect of lockdown on mental health problems through perceived social support was negative (B=−.01; β=−.01; *P*<.001), and the indirect effect of lockdown on mental health problems through maladaptive cognition was positive (B=.03, β=.04, *P*<.001). The indirect effects of quarantine on mental health problems through perceived social support (B=.02; β=.01; *P*<.001) and maladaptive cognition (B=.02; β=.01; *P*<.001) were positive. The total effects of lockdown (B=.03; β=.03; *P*=.001) and quarantine (B=.13; β=.05; *P*=.001) on mental health problems were statistically significant.

Moderation Test

The use of social media for obtaining health information during the COVID-19 pandemic significantly moderated the association between perceived social support and mental health problems ($\Delta\chi^2$=18.58, Δdf=2) (Multimedia Appendix 2). Specifically, the negative associations between perceived social support and mental health problems became stronger with the increased frequency of using social media (never: B=−.05; β=−.18; *P*<.001; sometimes: B=−.06; β=−.22; *P*<.001; always: B=−.08; β=−.27; *P*<.001).

The means of using prenatal care services during the COVID-19 pandemic significantly moderated the 3 model paths (Multimedia Appendix 3), including the paths from lockdown to maladaptive cognition ($\Delta\chi^2$=20.10, Δdf=3; went to hospital as usual: B=.16; β=.12; *P*<.001; made appointment with doctor: B=.07; β=.05; *P*<.001; online prenatal care: B=.13; β=.09; *P*=.03; uncertain: B=.05; β=.04; *P*=.25), the paths from mandatory quarantine to perceived social support ($\Delta\chi^2$=8.12, Δdf=3; went to hospital as usual: B=−.30; β=−.03; *P*=.004; made appointment with doctor: B=−.07; β=−.10; *P*=.504; online prenatal care: B=−.78; β=−.12; *P*=.01; uncertain: B=−.43; β=−.06; *P*=.07), and the paths from maladaptive cognition to mental health problems ($\Delta\chi^2$=11.48, Δdf=3; went to hospital as usual: B=.30; β=.41; *P*<.001; made appointment with doctor: B=.26; β=.37; *P*<.001; online prenatal care: B=.35; β=.42; *P*<.001; uncertain: B=.32; β=.38; *P*<.001).

Discussion

In this large-scale study, we report on the high prevalence of depression and anxiety in Chinese pregnant women during the COVID-19 pandemic and the potential sociodemographic, maternal, eHealth-related, control measure-related, cognitive, and social factors that affect them. Furthermore, we found that control measures were associated with depression and anxiety through social support and maladaptive cognition, and that the use of social media to obtain health information and the means of using prenatal care services were potential moderators of these associations.

Anxiety and depression are the most common mental disorders that occur during pregnancy, affecting between 10%-30% of pregnant women in China and other countries [58-61]. Our results suggest that the COVID-19 pandemic has resulted in a substantial increase in pregnant women’s risk of mental health problems, as 44.6% (8712/19,515) and 29.2% (5696/19,515) of our participants had probable mild to severe depression or anxiety, respectively. The significant background factors of anxiety/depression, including age, socioeconomic status, pregnancy-related status, and health service use, are in line with recent COVID-19 studies and non-COVID-19 studies [25,58,59,62]. In general, our results suggest that women with less social capital, experience in pregnancy, or health service resources/access experience more mental distress during the COVID-19 pandemic than those without such issues. In addition, this is the first study to reveal the associations between eHealth-related activities (ie, using online prenatal care services, making appointments with doctors, or using social media to obtain health information) and anxiety/depressive symptoms among pregnant women during the COVID-19 pandemic. We found that about half of the participants (10,189/19,515, 52.2%) went to hospital for prenatal care services as usual, while only 3.3% (640/19,515) used online prenatal care services. This may suggest that there is an urgent need to improve the quality of online prenatal care services and eHealth literacy and popularize the use of such services among pregnant women. These services allow pregnant women to access maternal health care with minimum COVID-19 exposure risk, which is desirable during the outbreak [63]. We found that 7.4% (1442/19,515) of pregnant women had self-harm/suicidal ideations in the past 2 weeks, which is slightly higher than the 5.2% prevalence rate reported before the COVID-19 epidemic [64]. Future studies should identify the causes of these mental health problems and the long-term impact of mental health problems on pregnant women. Tailored and timely interventions for mental health promotion are warranted, especially for vulnerable subgroups, such as pregnant women.

It is intriguing that in our study, lockdown and mandatory quarantine affected mental health problems in different ways through different underlying mechanisms. First, the total effect and indirect effects of lockdown on mental health problems were statistically significant, but the direct effects of lockdown were not. Lockdown increased the incidence of mental health problems through enhancing maladaptive cognition, which is consistent with previous studies on maladaptive cognition in the context of other stressful events (eg, daily hassles) [41,45]. Interestingly, we found that lockdown might increase perceived social support, which in turn might reduce the prevalence of mental health problems. However, a study in New Zealand found a nonsignificant change in perceived social support during the nationwide lockdown compared to that during the pre-COVID period [18]. The different associations between lockdown and perceived social support in the New Zealand study and our study may be partially due to the differences in study designs and samples, or the fact that the severity of lockdown measures and concomitant supporting measures varied across countries [1,2]. The positive association between lockdown and perceived social support in our study may be due to the fact that lockdown prohibits people from leaving an area, which might have increased participants' time and opportunities to stay and communicate with their families. A study conducted on April 2020 in Ireland (N=70) also reported that pregnant women improved their relationships with their partners by talking more, exercising together, and sharing tasks during lockdown [29]. In addition, governmental support has increased during the COVID-19 pandemic, and this has had a protective effect against anxiety among Chinese residents [65]. Thus, increased family support and governmental support during lockdown might explain why lockdown had a positive association with perceived social support and a negative indirect effect on mental health problems among Chinese pregnant women. These explanations should be assessed in future work, such as qualitative studies and case studies.

Second, as hypothesized, mandatory quarantine was significantly associated with a greater incidence of mental health problems. This result is consistent with several previous studies [19,20,66,67]. Furthermore, we found that reduced perceived social support and increased maladaptive cognition may explain this association, which is consistent with a study conducted by Zhu et al [21], who argued that the impacts of quarantine on daily life may explain the effect of quarantine on mental health problems. Quarantine is different from regional or nationwide lockdowns, as quarantine means that individuals are not allowed to leave the building or receive visitors. This difference may partially explain why lockdown increased perceived social support and mandatory quarantine reduced perceived social support in our study. Mandatory quarantine may also substantially affect other aspects of daily life, such as difficulties with quarantine compliance, inadequate information on prevention measures, decreased physical activity, the perceived high risk of COVID-19 infection [68], increased concerns over fetal safety, and difficulties in receiving prenatal care [23], which in turn aggravate mental health problems among pregnant women. Future studies should investigate these potential mediators and identify the cause of postquarantine changes in mental health status (eg, posttraumatic stress disorder development and increased stress).

Our findings highlight the importance of social and cognitive mechanisms in understanding the associations between lockdown/mandatory quarantine and mental health. In terms of our sample, the mediation effects of social and cognitive mechanisms accounted for large proportions of the total effects in the model. In general, the mediation model is supported by the conservation of resources theory [38] and the response styles theory [39,40], which explain how control measures for COVID-19 influence mental health. Based on these theories, future studies may explore other mediators, such as the loss and gain of other types of resources (eg, financial/personal resources) and other types of cognitive responses (eg, positive reappraisal). It is particularly important to identify modifiable psychosocial mediators because COVID-19 might become a persistent health threat, and such control measures might be inevitable [69]. Furthermore, our findings have important practical clinical and political implications. Since referrals are not feasible in the context of the COVID-19 pandemic, mental health first aid and brief non-face-to-face intervention services, such as the screening of mental distress in high-risk groups, counselling hotlines, and online education for problem-solving and stress-coping skills, should be made available for pregnant women and other vulnerable populations. In addition, it is important to guarantee that pregnant women who stay in areas under lockdown or mandatory quarantine can maintain regular communication with their significant others through social media, and receive adequate social support from both significant others and health/social care staff. More intensive therapies, such as cognitive behavioral therapy for cognitive restructuring and adaptive skill training, may be needed for those who experience great anxiety or depressive symptoms, such as the 29.2% (5696/19,515) and 44.6% (8712/19,515) of participants who had probable mild to severe anxiety or depression in our sample, respectively. Several environmental and structural factors, such as adequate preventive facilities, timely and accurate health information, access to multiple health service resources, and mass publicity for properly promoting accurate information on COVID-19, may also help pregnant women to reduce maladaptive cognition and facilitate positive reappraisal. Our results highlight the importance of integrating mental health care and eHealth into the implementation of control measures.

We also found that the use of social media for obtaining health information and the means of using prenatal care services during the COVID-19 pandemic were significant moderators in the model paths. Specifically, social media use strengthened the protective effect of perceived social support on mental health, and making appointments with doctors for prenatal care might buffer the adverse effects of lockdown and mandatory quarantine on maladaptive cognition and perceived social support, respectively. Furthermore, making such appointments might also buffer the adverse effects of maladaptive cognition on mental health. Using social media and making appointments with doctors may imply that people have various ways to access multiple health information and health service resources, which play an important role in buffering the stress appraisal and coping processes. These alternatives may be particularly useful for pregnant women during the COVID-19 pandemic, as high-quality and ongoing prenatal care is essential in supporting a healthy pregnancy and detecting risks early [70]. They are also helpful in reducing the risk of infection and related concerns among pregnant women [46]. In general, our results suggest that virtual visits and telemedicine should be included as part of a bundled care model.

Ours is one of the very few studies to assess pregnant women who experience and do not experience lockdown/mandatory quarantine and explore how such experiences might influence pregnant women’s interpersonal, cognitive, and mental statuses. Although our study adds to the literature on disease control and mental research, it has several limitations. First, the cross-sectional design prohibits causal inferences; the mediation model is exploratory and should not be interpreted as a causal mediation model. Our findings are intended for the generation of future research questions and provision of preliminary insights for when longitudinal studies are less feasible. Second, given the large sample size, the associations between small effect sizes could be statistically significant. Therefore, the interpretation of our results should be based on both statistical significance and effect size. Third, single items were used to assess perceived general social support and social support change; psychometric properties could not be established. Finally, the nonrandomly selected sample may have introduced selection bias, and the generalization of the study findings to other populations should be made cautiously. However, the large sample was recruited from all 34 provincial-level administrative regions in China. Therefore, our sample may accurately represent the population of China at a national level.

In conclusion, quarantine was strongly directly and indirectly associated with poor mental health status through decreased perceived social support and increased maladaptive cognition, while lockdown was indirectly associated with mental health through increased perceived social support and maladaptive cognition among pregnant women. The use of social media for obtaining health information and the means of using prenatal care service were significant moderators in the model paths. Follow-up studies are warranted to examine the long-term impacts of lockdown/quarantine control measures. The Chinese context of this study and the present global situation may differ in terms of the number of confirmed cases, the types and severity of control measures, and public responses toward COVID-19 and control measures. The validation of our findings and identification of similarities and differences across different countries are warranted.

Acknowledgments

We greatly appreciate Zhou Feng Rong from the Shandong Province Hospital for Women and Children's Health, Wu Ying Lan from the Hunan Provincial Maternal and Child Health Care Hospital, Zhao Juan from the Beijing Obstetrics and Gynecology Hospital Capital Medical University, Tian Hong from the Women and Children's Health Care Hospital of Liaoning Province, Qiu Li-qian from the Women's Hospital of the School of Medicine of the Zhejiang University, Xia Jianhong from the Guangdong Women and Children Hospital, Wang Lan from the Chongqing Health Center for Women and Children, and Li Fen from the First Affiliated Hospital of Medical College of Xi’an Jiaotong University for their assistance in data collection.

Authors' Contributions

XY conceived the research questions, conducted the statistical analysis, and drafted the manuscript. WQ and JTFL designed this study and assembled the team of collaborators. WQ supervised the project’s implementation. All authors assisted in questionnaire design, data collection, and data interpretation, and all authors provided comments with regard to the intellectual content of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Model path results in Amos.
[DOCX File , 19 KB - jmir_v23i1e24495_app1.docx]

Multimedia Appendix 2
Using social media for obtaining health information as a moderator of the model paths.
[PNG File , 85 KB - jmir_v23i1e24495_app2.png]

Multimedia Appendix 3
Means of using prenatal care services as a moderator of the model paths.
[PNG File , 89 KB - jmir_v23i1e24495_app3.png]

References

1. Wong SY, Kwok KO, Chan FKL. What can countries learn from Hong Kong's response to the COVID-19 pandemic? CMAJ 2020 May 11;192(19):E511-E515 [FREE Full text] [doi: 10.1503/cmaj.200563] [Medline: 32332040]
2. Quarantine and isolation. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/quarantine/index.html [accessed 2020-12-16]
3. Weekly epidemiological update - 21 September 2020. World Health Organization. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update---21-september-2020 [accessed 2020-12-16]
4. Dunford D, Dale B, Stylianou N, Lowther E, Ahmed M, de la Torre Arenas I. Coronavirus: The world in lockdown in maps and charts. BBC. 2020 Apr 06. URL: https://www.bbc.com/news/world-52103747 [accessed 2020-12-16]
5. Shi J. Beijing and Shanghai impose new controls on residents as China battles to contain coronavirus. South China Morning Post. 2020 Feb 10. URL: https://www.scmp.com/news/china/society/article/3049891/beijing-and-shanghai-impose-new-controls-residents-china-battles [accessed 2020-12-16]
6. Griffiths J, Woodyatt A. 780 million people in China are living under travel restrictions due to the coronavirus outbreak. CNN World. 2020 Feb 17. URL: https://www.cnn.com/2020/02/16/asia/coronavirus-covid-19-death-toll-update-intl-hnk/index.html [accessed 2020-12-16]
7. Considerations for quarantine of contacts of COVID-19 cases. World Health Organization. 2020 Aug 19. URL: https://www.who.int/publications/i/item/considerations-for-quarantine-of-individuals-in-the-context-of-containment-for-coronavirus-disease-(covid-19) [accessed 2020-12-16]
8. Protocol on Prevention and Control of COVID-19 (Edition 6). National Health Commission of the People's Republic of China. URL: http://english.www.gov.cn/2020special/5e32830ec6d019625c60433b/202003/26/content_WS5e7c683ec6d0c201c2cbf85c.html [accessed 2020-12-16]
9. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Apr 16;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
10. Governmental response to COVID-19. Korean Ministry of Health and Welfare. 2020 Feb 23. URL: https://www.mohw.go.kr/eng/nw/nw0101vw.jsp?PAR_MENU_ID=1007&MENU_ID=100701&page=1&CONT_SEQ=353125 [accessed 2020-12-30]
11. Self-isolation and treating coronavirus (COVID19) symptoms. National Health Service (England). URL: https://www.nhs.uk/conditions/coronavirus-covid-19/self-isolation-and-treatment/ [accessed 2020-12-30]
12. State Action on Coronavirus (COVID-19). National Conference of State Legislatures. URL: https://www.ncsl.org/research/health/state-action-on-coronavirus-covid-19.aspx [accessed 2020-12-16]
13. Feb 3: Daily briefing on novel coronavirus cases in China. National Health Commission of the People's Republic of China. URL: http://en.nhc.gov.cn/2020-02/03/c_76111.htm [accessed 2020-12-16]
14. National Health Commission of the People's Republic of China. Feb 12: Daily briefing on novel coronavirus cases in China. URL: http://en.nhc.gov.cn/2020-02/12/c_76463.htm [accessed 2020-12-16]
15. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
16. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
17. Concerns are raised over the threat of COVID-19 to mental health in Europe. United Nations. 2020 May 05. URL: https://unric.org/en/concerns-are-raised-over-the-threat-of-covid-19-to-mental-health-in-europe/ [accessed 2020-12-16]
18. Sibley CG, Greaves LM, Satherley N, Wilson MS, Overall NC, Lee CHJ, et al. Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. Am Psychol 2020;75(5):618-630. [doi: 10.1037/amp0000662] [Medline: 32496074]

19. Lei L, Huang X, Zhang S, Yang J, Yang L, Xu M. Comparison of Prevalence and Associated Factors of Anxiety and Depression Among People Affected by versus People Unaffected by Quarantine During the COVID-19 Epidemic in Southwestern China. Med Sci Monit 2020 Apr 26;26:e924609 [FREE Full text] [doi: 10.12659/MSM.924609] [Medline: 32335579]
20. Li J, Yang Z, Qiu H, Wang Y, Jian L, Ji J, et al. Anxiety and depression among general population in China at the peak of the COVID-19 epidemic. World Psychiatry 2020 Jun;19(2):249-250 [FREE Full text] [doi: 10.1002/wps.20758] [Medline: 32394560]
21. Zhu S, Wu Y, Zhu CY, Hong WC, Yu ZX, Chen ZK, et al. The immediate mental health impacts of the COVID-19 pandemic among people with or without quarantine managements. Brain Behav Immun 2020 Jul;87:56-58 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.045] [Medline: 32315758]
22. Zhang J, Lu H, Zeng H, Zhang S, Du Q, Jiang T, et al. The differential psychological distress of populations affected by the COVID-19 pandemic. Brain Behav Immun 2020 Jul;87:49-50 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.031] [Medline: 32304883]
23. Qiao J. What are the risks of COVID-19 infection in pregnant women? Lancet 2020 Mar 07;395(10226):760-762 [FREE Full text] [doi: 10.1016/S0140-6736(20)30365-2] [Medline: 32151334]
24. Ceulemans M, Hompes T, Foulon V. Mental health status of pregnant and breastfeeding women during the COVID-19 pandemic: A call for action. Int J Gynaecol Obstet 2020 Oct;151(1):146-147. [doi: 10.1002/ijgo.13295] [Medline: 32620037]
25. Wu Y, Zhang C, Liu H, Duan C, Li C, Fan J, et al. Perinatal depressive and anxiety symptoms of pregnant women during the coronavirus disease 2019 outbreak in China. Am J Obstet Gynecol 2020 Aug;223(2):240.e1-240.e9 [FREE Full text] [doi: 10.1016/j.ajog.2020.05.009] [Medline: 32437665]
26. Yue C, Liu C, Wang J, Zhang M, Wu H, Li C, et al. Association between social support and anxiety among pregnant women in the third trimester during the coronavirus disease 2019 (COVID-19) epidemic in Qingdao, China: The mediating effect of risk perception. Int J Soc Psychiatry 2020 Jul 09:20764020941567 [FREE Full text] [doi: 10.1177/0020764020941567] [Medline: 32643510]
27. Durankuş F, Aksu E. Effects of the COVID-19 pandemic on anxiety and depressive symptoms in pregnant women: a preliminary study. J Matern Fetal Neonatal Med 2020 May 18:1-7. [doi: 10.1080/14767058.2020.1763946] [Medline: 32419558]
28. Chasson M, Taubman-Ben-Ari O, Abu-Sharkia S. Jewish and Arab pregnant women's psychological distress during the COVID-19 pandemic: the contribution of personal resources. Ethn Health 2020 Sep 02:1-13. [doi: 10.1080/13557858.2020.1815000] [Medline: 32877202]
29. Milne SJ, Corbett GA, Hehir MP, Lindow SW, Mohan S, Reagu S, et al. Effects of isolation on mood and relationships in pregnant women during the covid-19 pandemic. Eur J Obstet Gynecol Reprod Biol 2020 Sep;252:610-611 [FREE Full text] [doi: 10.1016/j.ejogrb.2020.06.009] [Medline: 32616415]
30. Parra-Saavedra M, Villa-Villa I, Pérez-Olivo J, Guzman-Polania L, Galvis-Centurion P, Cumplido-Romero Á, et al. Attitudes and collateral psychological effects of COVID-19 in pregnant women in Colombia. Int J Gynaecol Obstet 2020 Nov;151(2):203-208. [doi: 10.1002/ijgo.13348] [Medline: 32799318]
31. Dong H, Hu R, Lu C, Huang D, Cui D, Huang G, et al. Investigation on the mental health status of pregnant women in China during the Pandemic of COVID-19. Arch Gynecol Obstet 2020 Oct 03:1-7 [FREE Full text] [doi: 10.1007/s00404-020-05805-x] [Medline: 33009997]
32. Zhou Y, Shi H, Liu Z, Peng S, Wang R, Qi L, et al. The prevalence of psychiatric symptoms of pregnant and non-pregnant women during the COVID-19 epidemic. Transl Psychiatry 2020 Sep 19;10(1):319 [FREE Full text] [doi: 10.1038/s41398-020-01006-x] [Medline: 32950999]
33. Zhang Y, Ma ZF. Psychological responses and lifestyle changes among pregnant women with respect to the early stages of COVID-19 pandemic. Int J Soc Psychiatry 2020 Aug 20:20764020952116 [FREE Full text] [doi: 10.1177/0020764020952116] [Medline: 32815434]
34. Patabendige M, Gamage MM, Weerasinghe M, Jayawardane A. Psychological impact of the COVID-19 pandemic among pregnant women in Sri Lanka. Int J Gynaecol Obstet 2020 Oct;151(1):150-153. [doi: 10.1002/ijgo.13335] [Medline: 32731307]
35. Taubman-Ben-Ari O, Chasson M, Sharkia SA, Weiss E. Distress and anxiety associated with COVID-19 among Jewish and Arab pregnant women in Israel. J Reprod Infant Psychol 2020 Jul;38(3):340-348. [doi: 10.1080/02646838.2020.1786037] [Medline: 32573258]
36. Ahorsu DK, Imani V, Lin CY, Timpka T, Broström A, Updegraff JA, et al. Associations Between Fear of COVID-19, Mental Health, and Preventive Behaviours Across Pregnant Women and Husbands: An Actor-Partner Interdependence Modelling. Int J Ment Health Addict 2020 Jun 11:1-15 [FREE Full text] [doi: 10.1007/s11469-020-00340-x] [Medline: 32837427]
37. Nanjundaswamy M, Shiva L, Desai G, Ganjekar S, Kishore T, Ram U, et al. COVID-19-related anxiety and concerns expressed by pregnant and postpartum women-a survey among obstetricians. Arch Womens Ment Health 2020 Aug 25:1-4 [FREE Full text] [doi: 10.1007/s00737-020-01060-w] [Medline: 32839898]

38. Hobfoll SE. Conservation of resources theory: Its implication for stress, health, and resilience. In: Folkman S, editor. The Oxford handbook of stress, health, and coping. Oxford, England: Oxford University Press; 2011.
39. Beck AT, Emery G, Greenberg RL. Anxiety disorders and phobias: A cognitive perspective. New York, New York, United States: Basic Books; 2005.
40. Nolen-Hoeksema S. Sex differences in unipolar depression: evidence and theory. Psychol Bull 1987 Mar;101(2):259-282. [doi: 10.1037/0033-2909.101.2.259] [Medline: 3562707]
41. Chan SM, Chan SK, Kwok WW. Ruminative and catastrophizing cognitive styles mediate the association between daily hassles and high anxiety in Hong Kong adolescents. Child Psychiatry Hum Dev 2015 Feb;46(1):57-66. [doi: 10.1007/s10578-014-0451-9] [Medline: 24570171]
42. Weems CF, Berman SL, Silverman WK, Saavedra LM. Cognitive errors in youth with anxiety disorders: The linkages between negative cognitive errors and anxious symptoms. Cognit Ther Res 2001;25(5):559-575.
43. Nolen-Hoeksema S. Responses to depression and their effects on the duration of depressive episodes. J Abnorm Psychol 1991 Nov;100(4):569-582. [doi: 10.1037//0021-843x.100.4.569] [Medline: 1757671]
44. Verstraeten K, Bijttebier P, Vasey M, Raes F. Specificity of worry and rumination in the development of anxiety and depressive symptoms in children. Br J Clin Psychol 2011 Nov;50(4):364-378. [doi: 10.1348/014466510X532715] [Medline: 22003947]
45. Michl LC, McLaughlin KA, Shepherd K, Nolen-Hoeksema S. Rumination as a mechanism linking stressful life events to symptoms of depression and anxiety: longitudinal evidence in early adolescents and adults. J Abnorm Psychol 2013 May;122(2):339-352 [FREE Full text] [doi: 10.1037/a0031994] [Medline: 23713497]
46. Khilnani A, Schulz J, Robinson L. The COVID-19 pandemic: new concerns and connections between eHealth and digital inequalities. Journal of Information, Communication and Ethics in Society 2020 Jun 25;18(3):393-403. [doi: 10.1108/JICES-04-2020-0052]
47. Hernández C, Valdera CJ, Cordero J, López E, Plaza J, Albi M. Impact of telemedicine on assisted reproduction treatment in the public health system. J Healthc Qual Res 2020;35(1):27-34. [doi: 10.1016/j.jhqr.2019.08.004] [Medline: 31883955]
48. Ybarra ML, Mitchell KJ, Palmer NA, Reisner SL. Online social support as a buffer against online and offline peer and sexual victimization among U.S. LGBT and non-LGBT youth. Child Abuse Negl 2015 Jan;39:123-136 [FREE Full text] [doi: 10.1016/j.chiabu.2014.08.006] [Medline: 25192961]
49. Kothgassner OD, Goreis A, Kafka JX, Kaufmann M, Atteneder K, Beutl L, et al. Virtual social support buffers stress response: An experimental comparison of real-life and virtual support prior to a social stressor. J Behav Ther Exp Psychiatry 2019 Jun;63:57-65. [doi: 10.1016/j.jbtep.2018.11.003] [Medline: 30454876]
50. Li J, Mo PKH, Wu AMS, Lau JTF. Roles of Self-Stigma, Social Support, and Positive and Negative Affects as Determinants of Depressive Symptoms Among HIV Infected Men who have Sex with Men in China. AIDS Behav 2017 Jan;21(1):261-273 [FREE Full text] [doi: 10.1007/s10461-016-1321-1] [Medline: 26896120]
51. Garnefski N, Kraaij V. Cognitive emotion regulation questionnaire – development of a short 18-item version (CERQ-short). Pers Individ Dif 2006 Oct;41(6):1045-1053. [doi: 10.1016/j.paid.2006.04.010]
52. Zhu X, Auerbach RP, Yao S, Abela JRZ, Xiao J, Tong X. Psychometric properties of the Cognitive Emotion Regulation Questionnaire: Chinese version. Cogn Emot 2008;22(2):288-307. [doi: 10.1080/02699930701369035]
53. Kroenke K, Spitzer RL. The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatr Ann 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
54. Tsai FJ, Huang YH, Liu HC, Huang KY, Huang YH, Liu SI. Patient health questionnaire for school-based depression screening among Chinese adolescents. Pediatrics 2014 Feb;133(2):e402-e409. [doi: 10.1542/peds.2013-0204] [Medline: 24446447]
55. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
56. Tong X, An D, McGonigal A, Park SP, Zhou D. Validation of the Generalized Anxiety Disorder-7 (GAD-7) among Chinese people with epilepsy. Epilepsy Res 2016 Feb;120:31-36. [doi: 10.1016/j.eplepsyres.2015.11.019] [Medline: 26709880]
57. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
58. Zhang Y, Muyiduli X, Wang S, Jiang W, Wu J, Li M, et al. Prevalence and relevant factors of anxiety and depression among pregnant women in a cohort study from south-east China. J Reprod Infant Psychol 2018 Nov;36(5):519-529. [doi: 10.1080/02646838.2018.1492098] [Medline: 30092662]
59. Ma X, Wang Y, Hu H, Tao XG, Zhang Y, Shi H. The impact of resilience on prenatal anxiety and depression among pregnant women in Shanghai. J Affect Disord 2019 May 01;250:57-64. [doi: 10.1016/j.jad.2019.02.058] [Medline: 30831542]
60. Goodman JH, Chenausky KL, Freeman MP. Anxiety disorders during pregnancy: a systematic review. J Clin Psychiatry 2014 Oct;75(10):e1153-e1184. [doi: 10.4088/JCP.14r09035] [Medline: 25373126]
61. Jones A. Help Seeking in the Perinatal Period: A Review of Barriers and Facilitators. Soc Work Public Health 2019;34(7):596-605. [doi: 10.1080/19371918.2019.1635947] [Medline: 31242074]
62. Liu X, Chen M, Wang Y, Sun L, Zhang J, Shi Y, et al. Prenatal anxiety and obstetric decisions among pregnant women in Wuhan and Chongqing during the COVID-19 outbreak: a cross-sectional study. BJOG 2020 Sep;127(10):1229-1240 [FREE Full text] [doi: 10.1111/1471-0528.16381] [Medline: 32583536]

63. Chen Y, Li Z, Zhang YY, Zhao WH, Yu ZY. Maternal health care management during the outbreak of coronavirus disease 2019. J Med Virol 2020 Jul;92(7):731-739. [doi: 10.1002/jmv.25787] [Medline: 32219871]
64. Shi P, Ren H, Li H, Dai Q. Maternal depression and suicide at immediate prenatal and early postpartum periods and psychosocial risk factors. Psychiatry Res 2018 Mar;261:298-306. [doi: 10.1016/j.psychres.2017.12.085] [Medline: 29331710]
65. Gan Y, Ma J, Wu J, Chen Y, Zhu H, Hall BJ. Immediate and delayed psychological effects of province-wide lockdown and personal quarantine during the COVID-19 outbreak in China. Psychol Med 2020 Aug 13:1-12 [FREE Full text] [doi: 10.1017/S0033291720003116] [Medline: 32787981]
66. Filgueiras A, Stults-Kolehmainen M. Factors linked to changes in mental health outcomes among Brazilians in quarantine due to COVID-19. medRxiv Preprint posted online on June 13, 2020. [doi: 10.1101/2020.05.12.20099374]
67. Rossi R, Socci V, Talevi D, Mensi S, Niolu C, Pacitti F, et al. COVID-19 pandemic and lockdown measures impact on mental health among the general population in Italy. An N=18147 web-based survey. medRxiv Preprint posted online on April 14, 2020. [doi: 10.1101/2020.04.09.20057802]
68. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
69. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020 May 22;368(6493):860-868 [FREE Full text] [doi: 10.1126/science.abb5793] [Medline: 32291278]
70. Cohen GJ, Committee on Psychosocial Aspects of Child and Family Health. The prenatal visit. Pediatrics 2009 Oct;124(4):1227-1232. [doi: 10.1542/peds.2009-1785] [Medline: 19786458]

Abbreviations

CERQ: Cognitive Emotion Regulation Questionnaire
CFI: comparative fit index
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
GAD-7: Generalized Anxiety Disorder 7-item
MIMIC: multiple indicators multiple causes
NNFI: nonnormed fit index
PHQ-9: Patient Health Questionnaire-9
PM: proportion of mediation
RMSEA: root mean square of error approximation

Edited by G Eysenbach; submitted 22.09.20; peer-reviewed by G Zhang, W Zhang; comments to author 16.10.20; revised version received 01.11.20; accepted 30.11.20; published 22.01.21.

Please cite as:
Yang X, Song B, Wu A, Mo PKH, Di J, Wang Q, Lau JTF, Wang L
Social, Cognitive, and eHealth Mechanisms of COVID-19–Related Lockdown and Mandatory Quarantine That Potentially Affect the Mental Health of Pregnant Women in China: Cross-Sectional Survey Study
J Med Internet Res 2021;23(1):e24495
URL: http://www.jmir.org/2021/1/e24495/
doi:10.2196/24495
PMID:33302251

Original Paper

Socioeconomic Disparities in Social Distancing During the COVID-19 Pandemic in the United States: Observational Study

Romain Garnier[1], PhD, DVM; Jan R Benetka[2], MS; John Kraemer[3], MPH, JD; Shweta Bansal[1], PhD

[1]Department of Biology, Georgetown University, Washington, DC, United States

[2]Unacast, Oslo, Norway

[3]Department of Health Systems Administration, Georgetown University, Washington, DC, United States

Corresponding Author:
Shweta Bansal, PhD
Department of Biology
Georgetown University
Reiss Science Building
37th and O Streets
Washington, DC, 20057
United States
Phone: 1 2026879256
Email: sb753@georgetown.edu

Abstract

Background: Eliminating disparities in the burden of COVID-19 requires equitable access to control measures across socio-economic groups. Limited research on socio-economic differences in mobility hampers our ability to understand whether inequalities in social distancing are occurring during the SARS-CoV-2 pandemic.

Objective: We aimed to assess how mobility patterns have varied across the United States during the COVID-19 pandemic and to identify associations with socioeconomic factors of populations.

Methods: We used anonymized mobility data from tens of millions of devices to measure the speed and depth of social distancing at the county level in the United States between February and May 2020, the period during which social distancing was widespread in this country. Using linear mixed models, we assessed the associations between social distancing and socioeconomic variables, including the proportion of people in the population below the poverty level, the proportion of Black people, the proportion of essential workers, and the population density.

Results: We found that the speed, depth, and duration of social distancing in the United States are heterogeneous. We particularly show that social distancing is slower and less intense in counties with higher proportions of people below the poverty level and essential workers; in contrast, we show that social distancing is intensely adopted in counties with higher population densities and larger Black populations.

Conclusions: Socioeconomic inequalities appear to be associated with the levels of adoption of social distancing, potentially resulting in wide-ranging differences in the impact of the COVID-19 pandemic in communities across the United States. These inequalities are likely to amplify existing health disparities and must be addressed to ensure the success of ongoing pandemic mitigation efforts.

(J Med Internet Res 2021;23(1):e24591) doi:10.2196/24591

KEYWORDS

COVID-19; SARS-CoV-2; disease ecology; nonpharmaceutical interventions; mobility data; economic; disparity; social distancing; equity; access; socioeconomic; infectious disease; mobility

Introduction

Treatment options and vaccines are being developed to address the COVID-19 pandemic [1]. However, while the ability to detect new infections remains limited by testing capacity and the lack of nationwide syndromic surveillance capabilities [2], nonpharmaceutical interventions represent the only immediate tools public health agencies can use to limit the size and spatial scale of the outbreak [3]. In the United States, state and local governments are primarily responsible for measures such as school or business closures [4]. Historically, similar measures have been used to respond to pandemics, including during plague outbreaks in the Middle Ages [5] and during the 1918 Spanish influenza pandemic [6]. Data collected during the early part of the ongoing COVID-19 pandemic, in particular on the dynamics of the outbreak in China, indicate that nonpharmaceutical interventions can be successful in limiting the size of COVID-19 outbreaks [7] and in delaying large-scale spread [3].

However, social distancing may be adopted differently across communities, especially in the United States, where workers in sectors such as transportation and food retail receive lower wages and represent a larger fraction of workers deemed essential than those in other sectors of the workforce [8,9]. Assessing this differential impact requires the use of fine-scale mobility data, a stream of information that has proven useful in the early assessment of social distancing measures in the United States [10], Italy [11], and France [12]. Digital technologies have taken center stage in the response to COVID-19 [13], and mobility data in particular have enabled assessment of the responses to nonpharmaceutical interventions [14,15]. Previous studies report large-scale reductions in movement, with numbers quickly reaching values typically observed during holiday periods [10]. Furthermore, in the United States, changes in mobility were associated with reductions in COVID-19 cases [16,17] and in the reproduction number of the disease [18], and mobility data also revealed that these changes were largely already underway when state or county stay-at-home orders were issued [16,19,20]. Previous studies using mobility data have also suggested potential inequalities in the ability to practice social distancing based on income [21], race and education [20], or the availability of health care providers [22]. However, most of these studies consider these determinants in isolation and do not allow to disentangle the potential additive effects of the socio-economic make-up of counties on the ability of their populations to practice social distancing. These studies have also largely focused on understanding how social distancing was influenced by local or regional decisions and whether socioeconomic factors changed the responses to state and local interventions.

Here, we focus on an ecological understanding of how mobility varies with socioeconomic characteristics rather than assessing what drives these changes. As outlined above, multiple studies have aimed to understand the causes of mobility behavior changes. However, we seek to understand how the patterns of mobility vary across socioeconomic characteristics (regardless of the cause) during different stages of the pandemic response. In particular, we ask how quickly, how deeply, and for how long mobility changes occurred in locations according to their socioeconomic characteristics. Our approach does not seek to differentiate between spontaneous changes in mobility, such as in response to news coverage, or changes in response to state- or county-mandated orders. Rather, we focus on the resulting changes in mobility and how these differ by location and socioeconomic status.

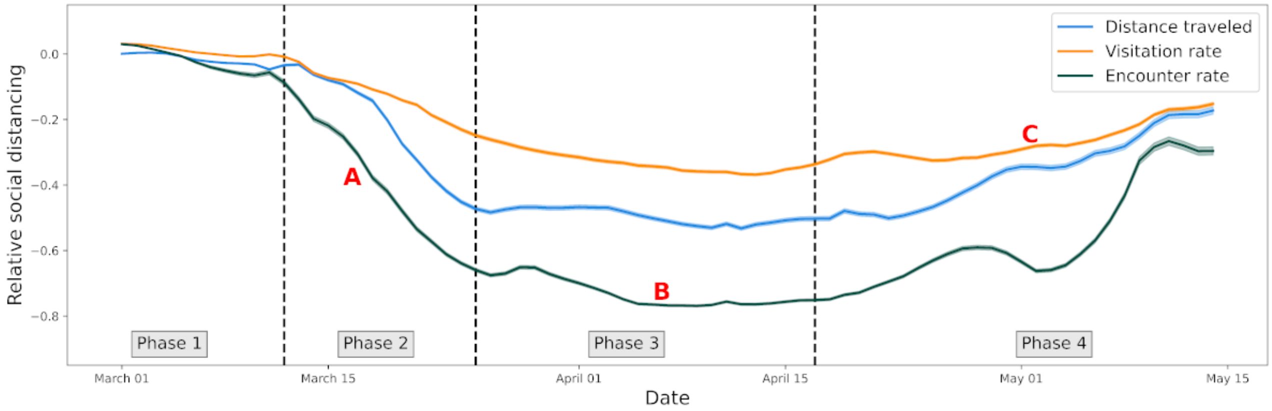
Methods

To measure mobility, we obtained daily county-specific mobility data for the United States from February 24 to May 14, 2020—the period during which most of the United States was simultaneously engaged in social distancing—through a partnership with Unacast [23]. The data set is based on the GPS location data collected from applications installed on tens of millions of devices, and it complies with the General Data Protection Regulation and the California Consumer Privacy Act [24]. The data set was shown to be representative by geographical location, income level, sex, and age in an analysis conducted by Unacast [25]. The fraction of all devices observed varies by location and time, and this has been captured in our analysis (details below).

As outcome measures, we considered data on the changes in three measures of mobility provided by Unacast: daily distance traveled (hereafter, “distance traveled”), rate of visitation to nonessential places (hereafter, “visitation rate”), and rate of encounters between devices within a 50-meter radius within an hour (hereafter, “encounter rate”). Distance traveled reflects the average distance between the home locations of users and locations visited in 1 day. Visitation rate reflects the number of visits to nonessential locations; the definition of nonessential venues is based on state-specific guidelines and policies and includes all locations other than those deemed essential (eg, food stores, pet stores, and pharmacies; more information can be found on Unacast’s website [26]). Encounter rate measures the likelihood of proximity between any two users within 50 meters over a one hour period. Each change is calculated relative to a county-specific baseline calculated from values obtained during a period of several weeks prior to the onset of major COVID-19–related changes in mobility in the United States (February 10 to March 4, 2020 for distance traveled and visitation rate; February 24 to March 4, 2020 for encounter rate). The resulting data set covers 3054 counties for the distance traveled and encounter rate and 2067 counties for the visitation rate. The county-level data on each measure described above can be accessed by contacting Unacast [23], and our model-processed data and the code used for the statistical analysis are available on GitHub [27].

To summarize the mobility time series, using the `fbprophet` package [28], we first fit a nonlinear model to the county-level social distancing time series, including a weekly trend to account for workweek variation. This package fits a piecewise regression while allowing setting the number of potential changepoints, and makes it possible to assess where breaks in the trend occur over the course of the social distancing time series. Most counties follow a dynamic similar to that of the mobility measures aggregated at the country level (Figure 1). In short, the underlying trend can be separated into four phases (ie, four breaks in the trend): phase 1, the baseline period; phase 2, the period of entry into social distancing, measured by the rate of mobility decrease; phase 3, the social distancing period, described by a sustained reduction in mobility outcomes; and phase 4, the period of exit from the social distancing phase, measured by the rate of mobility increase after sustained social distancing. This dynamic is specific to 2020; these four phases are not evident in 2019 (Figure S1 in Multimedia Appendix 1). From these model fits, we extracted several values that enabled us to characterize the changes in mobility during entry into and exit from the social distancing phase, as well as the mobility level during the period of sustained mobility reduction (see Figure 1 for details). Counties with no detectable increase in mobility after sustained social distancing (ie, where the slope of the trend at the end of the time series remains nonpositive) were not included in the phase 4 analysis. Each county time series was thus summarized by 3 values (or 2 values for the counties with no detectable phase 4 increase), which were used in the statistical analysis. We do not distinguish whether these changes are spontaneous in response to the COVID-19 pandemic or occur in response to public health policies; we only define the phases based on changes in mobility rates. That is, we do not seek to explain why the mobility changed but rather how it changed.

Figure 1. Time series of three mobility measures aggregated at the country level. We show a 7-day rolling mean of distance traveled (blue), visitation rate (orange), and encounters rate (green), with the solid line representing the mean and the shaded area two standard errors of the mean. The 4 phases in the mobility time series are delimited by dashed vertical lines in the figure and were generated from the model fit in each county independently. These phases allow us to calculate three summary measures for each time series: the slope of decline in phase 2 (A), the mean level of mobility during social distancing (B), and the slope of increase in mobility during exit from social distancing (C).



Our exposures of interest relate to the socioeconomic composition of each area: racial composition, population density, proportion living below the poverty level, and proportion of the workforce in industries designated as essential. Thus, our exposures of interest are area-level features, not individual features. We obtained information on racial composition and population density from the 2018 American Community Survey [29] and on the proportion of people below the poverty level from the Small Area Income and Poverty Estimate program [30]. We estimated the proportion of workers in industries designated as essential [31] from the Quarterly Census on Employment and Wages for the fourth quarter of 2019 [32].

We also adjusted for the fractions of devices observed in each county because sampling of mobile devices tends to vary geographically and over time.

We ran linear mixed models to analyze the associations between the social distancing summary values and the socioeconomic variables with state as a random effect using the standard 0.05 significance threshold. In the main analysis, we investigated independent associations with each covariate, and the resulting linear mixed model is of the form:

$$Y = X_i\beta_i + Z_\mu + \varepsilon$$

In our case, the response variable Y is one of the social distancing summary measures (the slope in phase 2, mean value in phase 3, or slope in phase 4). The socioeconomic predictors are included as the fixed effects X_i and the state as the random effect Z. The ε term captures the residuals. We ran independent models for each of the mobility measures (distance traveled, visitation rate, and encounter rate) and for each of the social distancing summary measures. We also performed a post hoc secondary analysis in which interaction terms between covariates were added to elucidate findings from the main analysis. Finally, as a sensitivity analysis, we fit an alternative linear model with state as a fixed effect rather than a random effect to adjust for any additional unmeasured state-level features. All analyses were conducted in Python 3.6.

The research presented in this paper was approved by the Georgetown-Medstar Institutional Review Board (study id STUDY00003041).

Results

Social distancing is heterogeneous at the county level (Figure 2). In counties with a higher proportion of people in poverty, social distancing was weaker: mobility was less restricted during the period of sustained mobility reduction, and the change occurred more slowly during the period of entry into social distancing for all three measures of mobility. Additionally, the resurgence in mobility was faster during the period of exit from sustained reductions in mobility for 2 of 3 measures of mobility (Figure 3). Counties with higher proportions of essential workers saw weaker social distancing adoption for all mobility markers during the period of sustained mobility reduction and a slower entry into social distancing based on two mobility markers. The rate of exit from the social distancing phase is less predicted by the proportion of essential workers. Contrastingly, in counties with a larger proportion of Black individuals or a higher population density, social distancing is stronger, with a faster entry into social distancing, lower mobility levels during the period of sustained mobility reduction, and a slower resurgence during the period of exit from the social distancing phase (Figure 3). The results were not significant for the encounter rate for the proportion of Black individuals in the population. Full statistical details are provided in Table S1 (Multimedia Appendix 1).

When interaction effects are added to the model of distance traveled in the period of entry into mobility reductions, all variables follow the same qualitative and quantitative patterns (Table S2, Multimedia Appendix 1). We found a significant positive interaction between the proportion of essential workers and the Black population, and we found a negative interaction between essential workers and low-income workers. The interaction between the proportions of Black people and low-income workers is not significantly associated with mobility.

Figure 2. Heterogeneity in mobility during social distancing. The map shows the average mobility during social distancing due to COVID-19 at the county level in the continental United States relative to the pre–COVID-19 baseline. A positive value indicates an increase in distance traveled, and a negative value indicates a decrease in distance traveled. Counties for which data are not available are shown in grey. The color map is centered at the 90% percentile of the decrease in mobility.

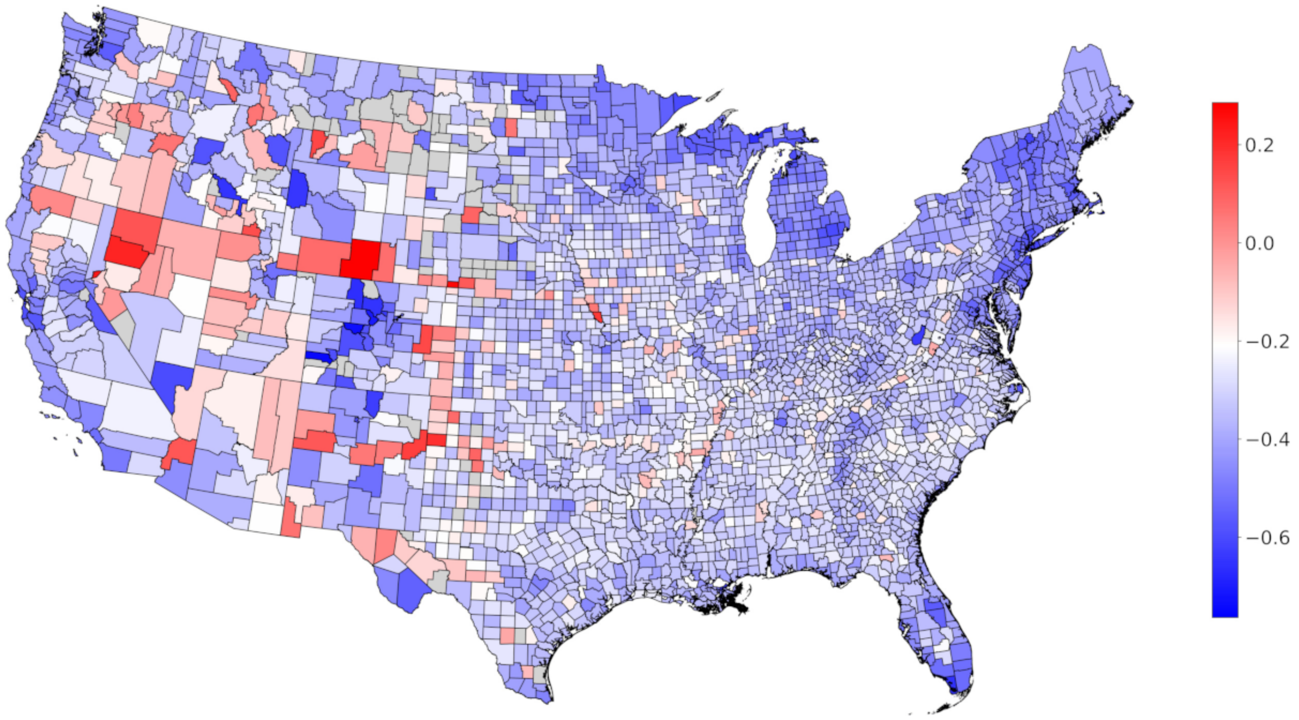
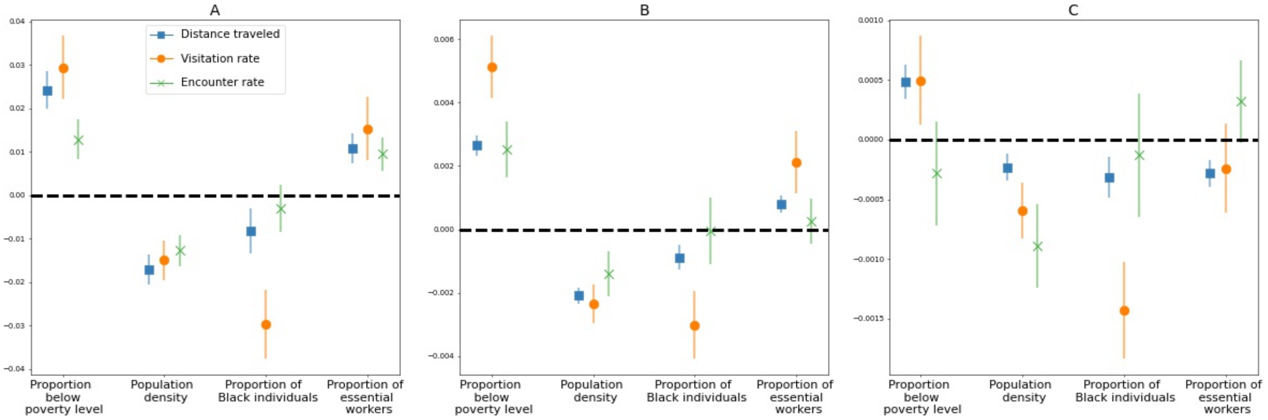


Figure 3. Regression coefficients of distance traveled, visitation rate, and encounter rate for the 4 socioeconomic factors associated with social distancing: (A) mobility during social distancing (phase 3); (B) decline in mobility during entry to the social distancing phase (phase 2); and (C) resurgence in mobility during exit from the social distancing phase (phase 4). The marker denotes the mean coefficient, and the error bars show the 95% confidence interval. A positive association (above the dashed line) indicates that an increase in a given factor leads to a weaker implementation of social distancing. A negative association (below the dashed line) indicates that an increase in the given factor is associated with a stronger implementation of social distancing measures.



Substituting the random effect state with a fixed effect yielded very similar results. There were only three differences of note: the association between the mobility resurgence measure as the visitation rate in the period of exit from the social distancing phase and the proportion of low-income workers became nonsignificant, while the associations between the resurgence in encounter rate in the period of exit from the social distancing phase and the proportions of essential and of low-wage workers became significant.

Discussion

The COVID-19 pandemic has highlighted significant health disparities in the United States, similar to the health inequities driven by income inequality and racial injustice that previously existed in the country [33]. Understanding the role of behavioral interventions in driving variations in the COVID-19 burden is crucial to our current and future outbreak response. Our study shows that changes in interaction in response to the pandemic are geographically heterogeneous and are associated with county-level socioeconomic factors. This is true for both the level of mobility restriction implemented during the social distancing phase and for the rate at which populations enter (“response engagement“) and exit (“response fatigue”) the social distancing phase.

Our analysis reveals that the occupational composition of the counties is associated with how deeply and for how long social distancing is maintained. Populations including more essential workers, who maintained food services, public transportation, and health care services during the pandemic [8], understandably participate less in social distancing and thus experience greater risk. Additionally, lower-income populations participate less in social distancing, likely in part because low-wage workers may have less access to job protections or paid leave. Our results provide further nuance to the analysis by Lou et al [9], who found that lower-wage workers were unable to reduce their work trips, in large part because businesses classified as essential tended to pay lower wages. Our results may help explain why lower-income counties have suffered a disproportionately high death burden from COVID-19 [17]; however, these results also need to be taken in light of the more general role that low income plays in negative health outcomes [21]. Reduced access to employer-sponsored health care [34] could further limit testing and treatment-seeking behavior and potentially worsen outbreaks in these communities. In rural communities (those with low population density in our study), the need to travel farther to access essential supplies and services such as food or health care [35] may also limit social distancing and would further confirm the existing disparity whereby rural counties suffer from poorer health outcomes than their more urban counterparts [36].

Importantly, we also found that counties with larger Black populations showed stronger adherence to social distancing measures during all phases, after controlling for the effects of income, occupation, and density. Our finding is supported by more local observations of differences between predominantly Black and White neighborhoods, such as Detroit [37]. We also found that social distancing remains more limited in populations that combine high proportions of Black individuals with high proportions of essential workers, possibly because minorities may be overrepresented in certain essential occupations [8]. Despite stronger distancing, there is growing evidence that African American communities experience higher rates of infection and death from COVID-19 [17,38,39]. We advocate for additional work on the structural racism that is at the root of these health disparities [40] and on the role of privilege in the differential burdens imposed by COVID-19 on a variety of communities [22,41].

There is a risk of the ecological fallacy if our results are interpreted as applying to individuals with the attributes we investigated rather than the share of attributes in communities. Survey and qualitative studies would help explain how individual, community, and public policy-level factors explain these associations.

Without large-scale test-trace-isolate programs or other interventions, intermittent social distancing will continue to be needed to contain cases and minimize the strain on health systems [42]. Technological solutions are being suggested and to an extent implemented [13-15]; however, these solutions are not without their limitations. The large-scale use of mobility data and other digital technologies (eg, for contact tracing) has opened up a debate on the responsible use of these emerging data streams [13,43], for instance to ensure that privacy concerns are properly assessed and addressed. These technologies would also likely be most effective with the implementation of a spatially and socially homogeneous testing strategy. Similarly, the long-term success and equity of a mitigation strategy hinges on paying more attention to the geographic heterogeneity in outbreak mitigation and focusing on the role of social and employment policies that affect the ability of individuals to engage in behavioral interventions.

Acknowledgments

The research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM123007. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The data and the code used for the statistical analysis are available at https://github.com/bansallab/SESdistancing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental material.
[DOCX File , 230 KB - jmir_v23i1e24591_app1.docx]

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
2. Dyer O. Covid-19: US testing ramps up as early response draws harsh criticism. BMJ 2020 Mar 23;368:m1167. [doi: 10.1136/bmj.m1167] [Medline: 32205342]
3. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
4. Gupta S, Nguyen T, Lozano RF, Raman S, Lee B, Bento A, et al. Tracking public and private response to the COVID-19 epidemicvidence from state and local government actions. NBER Working Paper Series. 2020 Apr. URL: http://www.nber.org/papers/w27027 [accessed 2020-09-25]
5. Gensini GF, Yacoub MH, Conti AA. The concept of quarantine in history: from plague to SARS. J Infect 2004 Nov;49(4):257-261 [FREE Full text] [doi: 10.1016/j.jinf.2004.03.002] [Medline: 15474622]
6. Markel H, Lipman HB, Navarro JA, Sloan A, Michalsen JR, Stern AM, et al. Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic. JAMA 2007 Aug 08;298(6):644-654. [doi: 10.1001/jama.298.6.644] [Medline: 17684187]
7. Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]
8. van Dorn A, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. Lancet 2020 Apr 18;395(10232):1243-1244 [FREE Full text] [doi: 10.1016/S0140-6736(20)30893-X] [Medline: 32305087]
9. Lou J, Shen X, Niemeier D. Are stay-at-home orders more difficult to follow for low-income groups? Journal of Transport Geography 2020 Dec;89:102894. [doi: 10.1016/j.jtrangeo.2020.102894]
10. Klein B, LaRock T, McCabe S, Torres L, Privitera F, Lake B, et al. Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak. Northeastern University Network Science Institute. 2020 Mar 31. URL: https://uploads-ssl.webflow.com/5c9104426f6f88ac129ef3d2/5e8374ee75221201609ab586_Assessing_mobility_changes_in_the_United_States_during_the_COVID_19_outbreak.pdf [accessed 2020-09-25]
11. Pepe E, Bajardi P, Gauvin L, Privitera F, Lake B, Cattuto C, et al. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. Sci Data 2020 Jul 08;7(1):230 [FREE Full text] [doi: 10.1038/s41597-020-00575-2] [Medline: 32641758]
12. Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. Population mobility reductions during COVID-19 epidemic in France under lockdown. medRxiv Preprint posted online on September 30, 2020. [doi: 10.1101/2020.05.29.20097097]
13. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
14. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv 2020 Jun;6(23):eabc0764 [FREE Full text] [doi: 10.1126/sciadv.abc0764] [Medline: 32548274]
15. Kishore N, Kiang MV, Engø-Monsen K, Vembar N, Schroeder A, Balsari S, et al. Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. Lancet Digit Health 2020 Nov;2(11):e622-e628 [FREE Full text] [doi: 10.1016/S2589-7500(20)30193-X] [Medline: 32905027]
16. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. Lancet Infect Dis 2020 Nov;20(11):1247-1254 [FREE Full text] [doi: 10.1016/S1473-3099(20)30553-3] [Medline: 32621869]
17. Abedi V, Olulana O, Avula V, Chaudhary D, Khan A, Shahjouei S, et al. Racial, Economic, and Health Inequality and COVID-19 Infection in the United States. J Racial Ethn Health Disparities 2020 Sep 01 [FREE Full text] [doi: 10.1007/s40615-020-00833-4] [Medline: 32875535]
18. Rubin D, Huang J, Fisher BT, Gasparrini A, Tam V, Song L, et al. Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States. JAMA Netw Open 2020 Jul 01;3(7):e2016099 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16099] [Medline: 32701162]
19. Siedner MJ, Harling G, Reynolds Z, Gilbert RF, Haneuse S, Venkataramani AS, et al. Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest-posttest comparison group study. PLoS Med 2020 Aug;17(8):e1003244 [FREE Full text] [doi: 10.1371/journal.pmed.1003244] [Medline: 32780772]
20. Huang V, Sutermaster S, Caplan Y, Kemp H, Schmutz D, Sgaier S. Social distancing across vulnerability, race, politics, and employment: How different Americans changed behaviors before and after major COVID-19 policy announcements Internet. medRxiv Preprint posted online on June 08, 2020. [doi: 10.1101/2020.06.04.20119131]

21. Weill JA, Stigler M, Deschenes O, Springborn MR. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. Proc Natl Acad Sci USA 2020 Aug 18;117(33):19658-19660. [doi: 10.1073/pnas.2009412117] [Medline: 32727905]
22. Dasgupta N, Jonsson Funk M, Lazard A, White B, Marshall S. Quantifying the Social Distancing Privilege Gap: A Longitudinal Study of Smartphone Movement. SSRN Journal Preprint posted online on August 11, 2020. [doi: 10.2139/ssrn.3588585]
23. The Social Distancing Scoreboard. Unacast. URL: https://www.unacast.com/covid19/social-distancing-scoreboard [accessed 2020-09-25]
24. Unacast. Archived at the Wayback Machine. URL: https://web.archive.org/web/20200916150140/https://www.unacast.com/privacy [accessed 2020-09-18]
25. How Accurate is Unacast Data? Unacast. Archived at the Wayback Machine. URL: https://web.archive.org/web/20200916225140/https://www.unacast.com/resources/how-accurate-is-unacast-data [accessed 2020-09-18]
26. Unacast Updates Social Distancing Scoreboard. Unacast. Archived at the Wayback Machine. URL: https://web.archive.org/web/20200916174247/https://www.unacast.com/post/unacast-updates-social-distancing-scoreboard [accessed 2020-09-18]
27. Garnier R, Benetka J, Kraemer J, Bansal S. SESdistancing. GitHub. URL: https://github.com/bansallab/SESdistancing [accessed 2020-09-18]
28. Taylor SJ, Letham B. Forecasting at Scale. Am Stat 2018 Apr 24;72(1):37-45. [doi: 10.1080/00031305.2017.1380080]
29. Annual county resident population estimates by age, sex, race, and hispanic origin: April 1, 2010 to July 1, 2018. US Census Bureau, Population Division. 2019 Jun. URL: https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2010-2018/cc-est2018-alldata.pdf?# [accessed 2020-09-25]
30. Small Area Income and Poverty Estimates (SAIPE). US Census Bureau. 2018. URL: https://www.census.gov/data-tools/demo/saipe/#/?map_geoSelector=aa_c [accessed 2020-09-25]
31. Tomer A, Kane J. How to protect essential workers during COVID-19. Brookings Reports. 2020 Mar 31. URL: https://www.brookings.edu/research/how-to-protect-essential-workers-during-covid-19/ [accessed 2020-09-25]
32. Quarterly Census of Employment and Wages. United States Bureau of Labor Statistics. URL: https://www.bls.gov/cew/ [accessed 2020-05-11]
33. Zimmerman FJ, Anderson NW. Trends in Health Equity in the United States by Race/Ethnicity, Sex, and Income, 1993-2017. JAMA Netw Open 2019 Jun 05;2(6):e196386 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.6386] [Medline: 31251377]
34. Gangopadhyaya A, Garrett AB. Unemployment, Health Insurance, and the COVID-19 Recession. SSRN Journal Preprint posted online on April 01, 2020. [doi: 10.2139/ssrn.3568489]
35. Bitto EA, Morton LW, Oakland MJ, Sand M. Grocery Store Acess Patterns in Rural Food Deserts. J Stud Food Soc 2015 Apr 27;6(2):35-48. [doi: 10.2752/152897903786769616]
36. Anderson TJ, Saman DM, Lipsky MS, Lutfiyya MN. A cross-sectional study on health differences between rural and non-rural U.S. counties using the County Health Rankings. BMC Health Serv Res 2015 Oct 01;15:441 [FREE Full text] [doi: 10.1186/s12913-015-1053-3] [Medline: 26423746]
37. Henry-Nickie M, Hudak J. Social distancing in Black and white neighborhoods in Detroit: A data-driven look at vulnerable communities Internet. Brookings. 2020 May 19. URL: https://www.brookings.edu/blog/fixgov/2020/05/19/social-distancing-in-black-and-white-neighborhoods-in-detroit-a-data-driven-look-at-vulnerable-communities/ [accessed 2020-09-17]
38. Yancy CW. COVID-19 and African Americans. JAMA 2020 May 19;323(19):1891-1892. [doi: 10.1001/jama.2020.6548] [Medline: 32293639]
39. Tai DBG, Shah A, Doubeni CA, Sia IG, Wieland ML. The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States. Clin Infect Dis 2020 Jun 20:ciaa815 [FREE Full text] [doi: 10.1093/cid/ciaa815] [Medline: 32562416]
40. Chowkwanyun M, Reed AL. Racial Health Disparities and Covid-19 - Caution and Context. N Engl J Med 2020 Jul 16;383(3):201-203. [doi: 10.1056/NEJMp2012910] [Medline: 32374952]
41. Brown EA, White BM. Recognizing Privilege as a Social Determinant of Health During COVID-19. Health Equity 2020;4(1):1-4 [FREE Full text] [doi: 10.1089/heq.2020.0038] [Medline: 32908954]
42. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020 May 22;368(6493):860-868 [FREE Full text] [doi: 10.1126/science.abb5793] [Medline: 32291278]
43. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]

Edited by M Focsa; submitted 25.09.20; peer-reviewed by C Abbafati, S Gao; comments to author 27.11.20; revised version received 10.12.20; accepted 14.12.20; published 22.01.21.

Please cite as:
Garnier R, Benetka JR, Kraemer J, Bansal S
Socioeconomic Disparities in Social Distancing During the COVID-19 Pandemic in the United States: Observational Study
J Med Internet Res 2021;23(1):e24591
URL: http://www.jmir.org/2021/1/e24591/
doi:10.2196/24591
PMID:33351774

Original Paper

Toward Using Twitter for Tracking COVID-19: A Natural Language Processing Pipeline and Exploratory Data Set

Ari Z Klein[1], PhD; Arjun Magge[1], PhD; Karen O'Connor[1], MS; Jesus Ivan Flores Amaro[1], BS; Davy Weissenbacher[1], PhD; Graciela Gonzalez Hernandez[1], PhD

Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Ari Z Klein, PhD
Department of Biostatistics, Epidemiology, and Informatics
Perelman School of Medicine
University of Pennsylvania
421A Blockley Hall
423 Guardian Dr
Philadelphia, PA, 19104
United States
Phone: 1 215 746 1101
Email: ariklein@pennmedicine.upenn.edu

Abstract

Background: In the United States, the rapidly evolving COVID-19 outbreak, the shortage of available testing, and the delay of test results present challenges for actively monitoring its spread based on testing alone.

Objective: The objective of this study was to develop, evaluate, and deploy an automatic natural language processing pipeline to collect user-generated Twitter data as a complementary resource for identifying potential cases of COVID-19 in the United States that are not based on testing and, thus, may not have been reported to the Centers for Disease Control and Prevention.

Methods: Beginning January 23, 2020, we collected English tweets from the Twitter Streaming application programming interface that mention keywords related to COVID-19. We applied handwritten regular expressions to identify tweets indicating that the user potentially has been exposed to COVID-19. We automatically filtered out “reported speech” (eg, quotations, news headlines) from the tweets that matched the regular expressions, and two annotators annotated a random sample of 8976 tweets that are geo-tagged or have profile location metadata, distinguishing tweets that self-report potential cases of COVID-19 from those that do not. We used the annotated tweets to train and evaluate deep neural network classifiers based on bidirectional encoder representations from transformers (BERT). Finally, we deployed the automatic pipeline on more than 85 million unlabeled tweets that were continuously collected between March 1 and August 21, 2020.

Results: Interannotator agreement, based on dual annotations for 3644 (41%) of the 8976 tweets, was 0.77 (Cohen κ). A deep neural network classifier, based on a BERT model that was pretrained on tweets related to COVID-19, achieved an F_1-score of 0.76 (precision=0.76, recall=0.76) for detecting tweets that self-report potential cases of COVID-19. Upon deploying our automatic pipeline, we identified 13,714 tweets that self-report potential cases of COVID-19 and have US state–level geolocations.

Conclusions: We have made the 13,714 tweets identified in this study, along with each tweet’s time stamp and US state–level geolocation, publicly available to download. This data set presents the opportunity for future work to assess the utility of Twitter data as a complementary resource for tracking the spread of COVID-19.

(J Med Internet Res 2021;23(1):e25314) doi:10.2196/25314

KEYWORDS

natural language processing; social media; data mining; COVID-19; coronavirus; pandemics; epidemiology; infodemiology

Introduction

In the United States, the rapidly evolving COVID-19 outbreak, the shortage of available testing, and the delay of test results have presented challenges for actively monitoring the spread of COVID-19 based on testing alone. An approach that has emerged for detecting cases without the need for extensive testing relies on voluntary self-reports of symptoms from the general population [1]. Considering that nearly one of every four adults in the United States already uses Twitter, and nearly half of them use it on a daily basis [2], researchers have begun exploring tweets for mentions of COVID-19 symptoms [3-8]. However, considering the incubation period of COVID-19 [9], detecting cases based on symptoms may not maximize the potential of Twitter data for real-time monitoring. The objective of this study was to develop, evaluate, and deploy a natural language processing (NLP) pipeline that automatically collects tweets reporting personal information more broadly—that is, beyond symptoms—that might indicate exposure to COVID-19 in the United States. In this paper, we present a publicly available data set containing 13,714 tweets that were identified by our automatic NLP pipeline between March 1 and August 21, 2020, with each tweet's time stamp and US state–level geolocation. This data set presents the opportunity to explore the use of Twitter data as a complementary resource "to understand and model the transmission and trajectory of COVID-19" [10].

Methods

Data Collection and Annotation

The Institutional Review Board (IRB) of the University of Pennsylvania reviewed this study and deemed it to be exempt human subjects research under Category (4) of Paragraph (b) of the US Code of Federal Regulations Title 45 Section 46.101 for publicly available data sources (45 CFR §46.101(b)(4)).

Between January 23 and March 20, 2020, we collected more than 7 million publicly available tweets that mention keywords related to COVID-19, are posted in English, are not retweets, and are geo-tagged or have user profile location metadata. We developed handwritten regular expressions (Multimedia Appendix 1)—search patterns designed to automatically match text strings—to identify a subset of the 7 million tweets that indicate that the user potentially has been exposed to COVID-19. Our query patterns were designed primarily to help identify potential cases of COVID-19 that are not based on testing and, thus, may not have been reported to the Centers for Disease Control and Prevention (CDC) [11]. The regular expressions matched approximately 160,000 (2%) of the 7 million tweets. Approximately 30,000 (19%) of the 160,000 matching tweets were then automatically removed using a system we developed in recent work [12] for filtering out "reported speech" (eg, quotations, news headlines) from health-related social media data.

In preliminary work [13], two annotators annotated a random sample of 10,000 of the 130,000 filtered tweets, and annotation guidelines (Multimedia Appendix 2) were developed to help the annotators distinguish between three classes of tweets. However, since then, we have removed 1024 of the annotated tweets that were collected from the Twitter Streaming application programming interface (API) based on a keyword that we have stopped using, and we have unified two of the classes. "Potential case" tweets include those that indicate that the user or a member of the user's household was denied testing for COVID-19, showing symptoms of COVID-19, potentially exposed to presumptive or confirmed cases of COVID-19, or had had experiences that pose a higher risk of exposure to COVID-19. "Other" tweets are related to COVID-19 and may discuss topics such as testing, symptoms, traveling, or social distancing, but do not indicate that the user or a member of the user's household may be infected. Among the 8976 tweets, 3644 (41%) were annotated by both annotators. Upon resolving the annotators' disagreements, 1456 (16%) of the tweets were annotated as "potential case" and 7520 (84%) as "other." Textbox 1 presents (slightly modified) sample tweets that match our handwritten regular expressions and were manually annotated as "potential case."

Textbox 1. Sample (slightly modified) tweets that match our handwritten regular expressions and were manually annotated as potential cases of COVID-19.

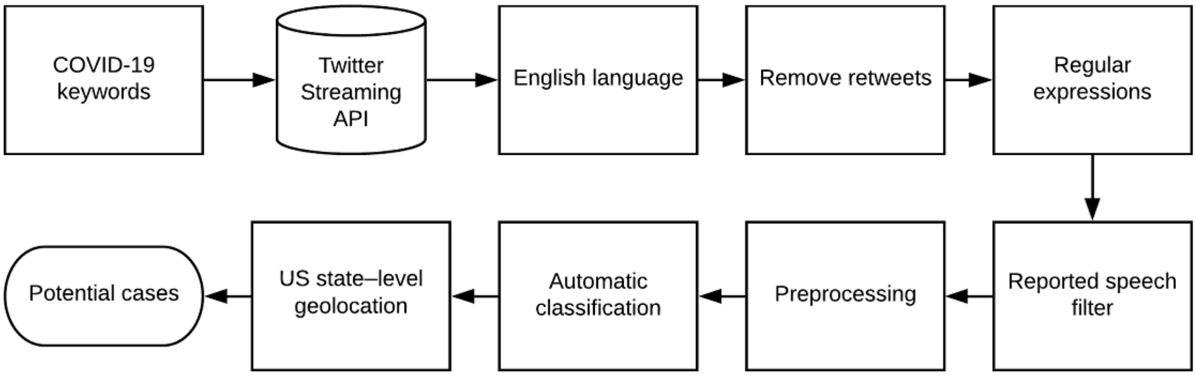
1. Nearly two weeks ago I had a fever, sore throat, runny nose, and cough. I want to know if it was coronavirus or just the common cold
2. My coworker in next office probably has #coronavirus. He and his wife have the symptoms, but they went to the hospital to get tested and were refused.
3. This girl in my class had the coronavirus, so I'm making an appointment with my doctor for a check up
4. Pretty sure I had a patient tonight with Coronavirus. Had all the symptoms and tested negative for the flu.
5. Why can celebrities, sports athletes & politicians without symptoms get tested, but my symptomatic child who has a compromised immune system cannot? #coronavirus
6. Since getting back from Seattle I've been sick and want to get a #coronavirus check. Called my PCP, they said to call health dept. Called them, they said I need to go thru my PCP. Called my PCP again, they said they can't help me
7. I'm convinced I have coronavirus. I've been to NYC, Phoenix, and San Diego in the last few weeks. I have a cough, a runny nose, and I'm really hot #covid19
8. Scared of the coronavirus because I have a sore throat and a headache I think its just a cold but I take the tube 4 times a day
9. Can't even get testing SCHEDULED while self-quarantined (my decision) and having coronavirus symptoms I take train thru New Rochelle to Manhattan
10. I have a bad cold. I went to the doctor, got some medications, the norm. But they couldn't rule out coronavirus because they don't have the tests.

As Textbox 1 illustrates, our handwritten regular expressions are based on query patterns designed to identify tweets that report personal information that may be useful for tracking potential cases of COVID-19, including not only symptoms (tweet 1), but also exposure to potential cases and a lack of access to COVID-19 testing. For example, our regular expressions retrieve tweets reporting that the user may have come in contact with coworkers (tweet 2), classmates (tweet 3), patients (tweet 4), and family members (tweet 5) who may have COVID-19, and potential exposure to COVID-19 through traveling (tweets 6 and 7) and commutes (tweets 8 and 9). Our regular expressions also retrieve tweets reporting that the user (tweet 9 and 10), a family member (tweet 5), or someone else that the user has been in contact with (tweet 2) was denied access to testing, even though they are sick. Since none of the tweets in Textbox 1 report being tested for or diagnosed with COVID-19, they represent potential cases that may not have been reported to the CDC.

Automatic Classification and Geolocation

We split the 8976 annotated tweets into 80% (7181 tweets) and 20% (1795 tweets) random sets—a training set (Multimedia Appendix 3) and held-out test set, respectively—for automatic classification. We used the *ktrain* [14] Python library to train and evaluate two supervised deep neural network classifiers based on bidirectional encoder representations from transformers (BERT): BERT-Base-Uncased [15] and COVID-Twitter-BERT [16]. After feeding the sequence of tweet tokens to BERT, the encoded representation is passed to a dropout layer (dropping rate of 0.1), followed by a dense layer with 2 units and a softmax activation, which predicts the class for each tweet. For training, we used Adam optimization with rate decay and warm-up. We used a batch size of 64, training runs for 3 epochs, and a maximum learning rate of 1×10^{-5}. We fine-tuned all layers of the transformer model with our annotated tweets. Prior to automatic classification, we preprocessed the tweets by normalizing usernames and URLs, and lowercasing the text. Figure 1 illustrates our automatic NLP pipeline for detecting tweets that indicate potential cases of COVID-19 in the United States. We deployed the pipeline on more than 85 million unlabeled tweets that were continuously collected between March 1 and August 21, 2020. We used Carmen [17] to infer the geolocation—at the US state level—of tweets that the classifier predicted as potential cases.

Figure 1. Automatic natural language processing (NLP) pipeline for detecting tweets that self-report potential cases of COVID-19 in the United States.

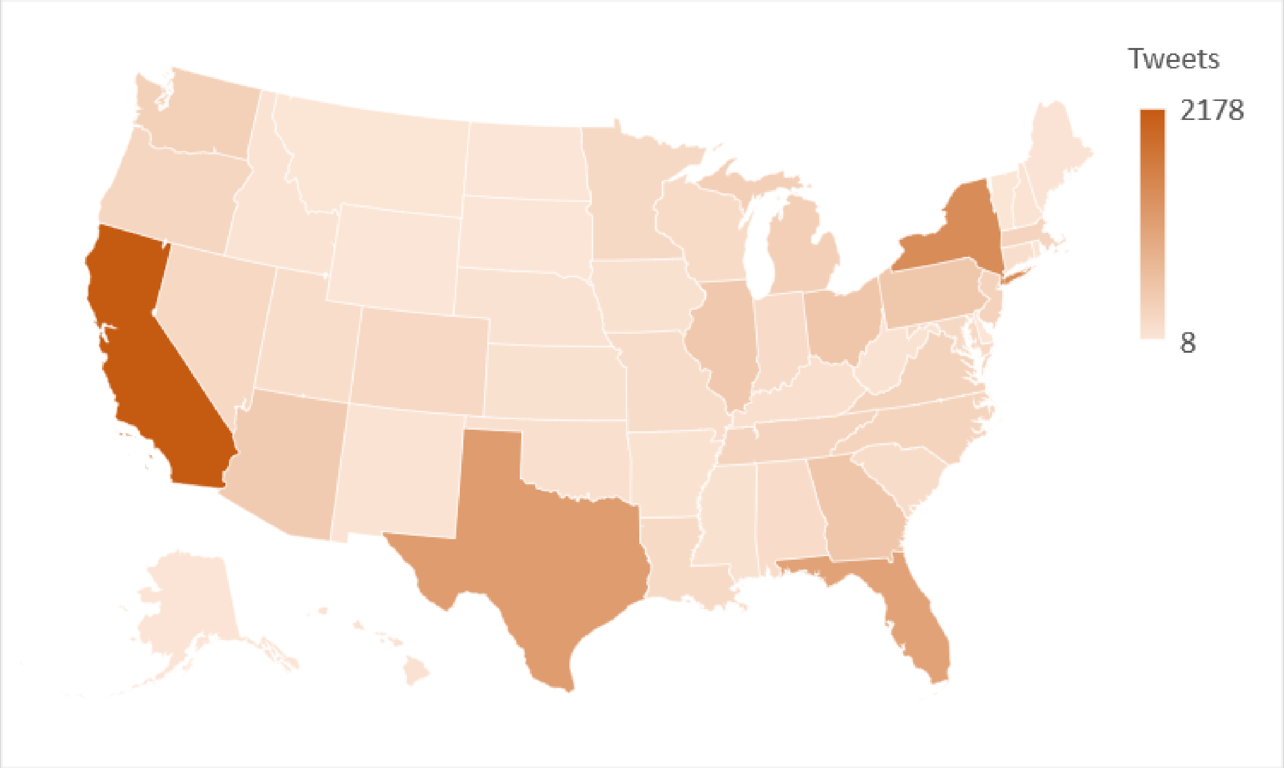


Results

Interannotator agreement, based on dual annotations for 3644 (41%) of the 8976 tweets, was 0.77 (Cohen κ), considered "substantial agreement" [18]. We evaluated two deep neural network classifiers on a held-out test set of 1795 (20%) of the 8976 tweets. The classifier based on the BERT-Base-Uncased pretrained model achieved an F_1-score of 0.70 (precision=0.72, recall=0.67) for the "potential case" class, and the classifier based on the COVID-Twitter-BERT pretrained model achieved an F_1-score of 0.76 (precision=0.76, recall=0.76), where:

We deployed our automatic pipeline, using the COVID-Twitter-BERT classifier, on more than 85 million unlabeled tweets that were continuously collected from the Twitter Streaming API between March 1 and August 21, 2020. Among the subset of tweets that were posted in English, not retweets, matched the regular expressions, and were not filtered out as reported speech, the COVID-Twitter-BERT classifier detected 13,714 "potential case" tweets for which Carmen inferred a US state–level geolocation. Figure 2 illustrates the ranges of "potential case" tweets that were automatically detected per state. We automatically detected "potential case" tweets from all 50 states, with the highest numbers posted in California, New York, Texas, and Florida.

Figure 2. Tweets self-reporting potential cases of COVID-19 in the United States, by state, between March 1 and August 21, 2020.



Discussion

Principal Findings

While Twitter data has been used to identify self-reports of symptoms by people who have tested positive for COVID-19 [3,4], the shortage of available testing and the delay of test results in the United States motivated us to assess whether Twitter data could be scaled to identify potential cases of COVID-19 that are not based on testing and, thus, may not have been reported to the CDC. There are studies that have not limited their exploration of COVID-19 symptoms on Twitter to users who have tested positive for COVID-19 [5-8]; however, limiting the detection of potential cases to symptoms may still underutilize the information available on Twitter. Our automatic NLP pipeline has detected potential cases of COVID-19 across the entire United States that are neither based on testing nor limited to symptoms, providing the opportunity to explore the utility of Twitter data more broadly as a complementary resource for tracking the spread of COVID-19. An analysis based on this data set is beyond the scope of this study. The 13,714 “potential case” tweets identified in this study can be downloaded using a Python script [19] and the input file in Multimedia Appendix 4, which contains the user ID, tweet ID, time stamp, and inferred state-level geolocation for each tweet. The script downloads the tweets that are still publicly available.

Conclusions

This paper presented an automatic NLP pipeline that was used to identify 13,714 tweets self-reporting potential cases of COVID-19 in the United States between March 1 and August 21, 2020, that may not have been reported to the CDC. This publicly available data set presents the opportunity for future work to assess the utility of Twitter data as a complementary resource for tracking the spread of COVID-19.

Acknowledgments

AZK contributed to the methodology, formal analysis, investigation, data curation, and writing the original draft. AM contributed to the software development, formal analysis, investigation, and writing the original draft. KO contributed to the data curation and writing (review and editing). JIFA contributed to the software development and writing (review and editing). DW contributed to the software development, formal analysis, investigation, and writing (review and editing). GGH contributed to the conceptualization, writing (review and editing), supervision, and funding acquisition. The authors would like to thank Alexis Upshur for contributing to annotating the Twitter data. This work was supported by the National Institutes of Health (NIH) National Library of Medicine (NLM; grant number R01LM011176) and National Institute of Allergy and Infectious Diseases (NIAID; grant number R01AI117011).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Regular expressions.
[TXT File , 3 KB - jmir_v23i1e25314_app1.txt]

Multimedia Appendix 2
Annotation guidelines.
[PDF File (Adobe PDF File), 1060 KB - jmir_v23i1e25314_app2.pdf]

Multimedia Appendix 3
Training data.
[TXT File , 249 KB - jmir_v23i1e25314_app3.txt]

Multimedia Appendix 4
Exploratory Twitter data set.
[TXT File , 851 KB - jmir_v23i1e25314_app4.txt]

References

1. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
2. Smith A, Anderson M. Social media use in 2018. Pew Research Center. 2018 Mar 01. URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2020-09-29]
3. Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi M, Yang Y. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. J Am Med Inform Assoc 2020 Aug 01;27(8):1310-1315. [doi: 10.1093/jamia/ocaa116] [Medline: 32620975]
4. Jeon J, Baruah G, Sarabadani S, Palanica A. Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data. J Med Internet Res 2020 Oct 02;22(10):e20509 [FREE Full text] [doi: 10.2196/20509] [Medline: 32936770]
5. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]
6. Panuganti BA, Jafari A, MacDonald B, DeConde AS. Predicting COVID-19 Incidence Using Anosmia and Other COVID-19 Symptomatology: Preliminary Analysis Using Google and Twitter. Otolaryngol Head Neck Surg 2020 Sep;163(3):491-497 [FREE Full text] [doi: 10.1177/0194599820932128] [Medline: 32484425]
7. Guntuku SC, Sherman G, Stokes DC, Agarwal AK, Seltzer E, Merchant RM, et al. Tracking Mental Health and Symptom Mentions on Twitter During COVID-19. J Gen Intern Med 2020 Sep 07;35(9):2798-2800 [FREE Full text] [doi: 10.1007/s11606-020-05988-8] [Medline: 32638321]
8. Guo J, Radloff CL, Wawrzynski SE, Cloyes KG. Mining twitter to explore the emergence of COVID-19 symptoms. Public Health Nurs 2020 Nov 16;37(6):934-940. [doi: 10.1111/phn.12809] [Medline: 32937679]
9. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine 2020 May 05;172(9):577-582. [doi: 10.7326/m20-0504]
10. Merchant RM, Lurie N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. JAMA 2020 May 26;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
11. United States COVID-19 cases and deaths by state. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/#cases_totalcases [accessed 2020-09-29]
12. Klein AZ, Cai H, Weissenbacher D, Levine LD, Gonzalez-Hernandez G. A natural language processing pipeline to advance the use of Twitter data for digital epidemiology of adverse pregnancy outcomes. Journal of Biomedical Informatics: X 2020 Dec;8:100076. [doi: 10.1016/j.yjbinx.2020.100076]
13. Klein A, Magge A, O'Connor K, Cai H, Weissenbacher D, Gonzalez-Hernandez G. A chronological and geographical analysis of personal reports of COVID-19 on Twitter. medRxiv Preprint published online on April 24, 2020. [doi: 10.1101/2020.04.19.20069948]
14. Maiya AS. ktrain: a low-code library for augmented machine learning. arXiv Preprint posted online on April 19, 2020. [FREE Full text]
15. Devlin J, Cheng M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT); June 2-7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]

16. Müller M, Salathé M, Kummervold P. COVID-Twitter-BERT: a natural language processing model to analyse COVID-19 content on Twitter. arXiv Preprint posted online on May 15, 2020. [FREE Full text]
17. Drezde M, Paul M, Bergsma S, Tran H. Carmen: a Twitter geo-location system with applications to public health. 2013 Presented at: Association for the Advancement of Artificial Intelligence (AIII) 2013 Workshop Expanding the Boundaries of Health Informatics Using AI; July 14-15, 2013; Bellevue, WA, USA.
18. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
19. Study data set. Bitbucket. URL: https://bitbucket.org/pennhlp/twitter_data_download/src/master/ [accessed 2021-01-18]

Abbreviations

API: application programming interface
BERT: bidirectional encoder representations from transformers
CDC: Centers for Disease Control and Prevention
NLP: natural language processing

Edited by G Eysenbach; submitted 27.10.20; peer-reviewed by K Verspoor, V Foufi, X Ji, L Sheets; comments to author 05.12.20; revised version received 14.12.20; accepted 14.12.20; published 22.01.21.

Please cite as:
Klein AZ, Magge A, O'Connor K, Flores Amaro JI, Weissenbacher D, Gonzalez Hernandez G
Toward Using Twitter for Tracking COVID-19: A Natural Language Processing Pipeline and Exploratory Data Set
J Med Internet Res 2021;23(1):e25314
URL: http://www.jmir.org/2021/1/e25314/
doi:10.2196/25314
PMID:33449904

Original Paper

Adequacy of Web-Based Activities as a Substitute for In-Person Activities for Older Persons During the COVID-19 Pandemic: Survey Study

Jiska Cohen-Mansfield[1,2,3], PhD; Aline Muff[2], PhD; Guy Meschiany[2], MA; Shahar Lev-Ari[1], PhD

[1]Department of Health Promotion, School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[2]Minerva Center for the Interdisciplinary Study of End of Life, Tel Aviv University, Tel Aviv, Israel
[3]Igor Orenstein Chair for the Study of Geriatrics, Tel Aviv University, Tel Aviv, Israel

Corresponding Author:
Jiska Cohen-Mansfield, PhD
Department of Health Promotion
School of Public Health, Sackler Faculty of Medicine
Tel Aviv University
POB 39040
Tel Aviv
Israel
Phone: 972 03 6407576
Email: jiska@tauex.tau.ac.il

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/2/e27687/

Abstract

Background: Senior centers and other types of clubs provide activities for older adults to address boredom, social isolation, and loneliness. Due to the COVID-19 pandemic, most of these activities have been cancelled. A limited range of web-based activities have been offered as alternatives. However, the effectiveness of these web-based group activities for older adults has scarcely been researched.

Objective: We aimed to understand the extent to which web-based activities for older adults provide an adequate substitute for in-person activities.

Methods: In this telephone survey, we interviewed 105 older adults in Israel who had been offered the opportunity to participate in web-based activities after routine activities closed due to the COVID-19 pandemic. Of the total sample, 49/105 (46.7%) participated in the activities and 56/105 (53.3%) did not. We inquired about the respondents' background characteristics, satisfaction with the activities, and reasons for participation or nonparticipation.

Results: The respondents who participated in the web-based activities tended to be highly satisfied with at least some of them. They rated the enjoyment derived from the content of the activity as the most important motivator, followed by maintaining a routine and by enjoying the group and the presence of others. Over 50% of the participants (28/49, 57%) wished to continue with the exercise programming after the end of the COVID-19 pandemic, and 41% (20/49) wished to continue with the web-based lectures. Participants were more likely to report partaking in alternative activities than nonparticipants (*P*=.04). The most common reasons cited by nonparticipants were being unaware of the web-based program (24/56, 43%) despite a notification having been sent to the entire sample, lack of interest in the content (18/56, 32%), and technical issues (13/56, 23%), such as not owning or being able to fully use a computer. Both participants and nonparticipants were interested in a wide range of topics, with many being very particular about the topics they wished to access. Approximately half expressed willingness to pay for access; those who were willing to pay tended to have more years of education (*P*=.03).

Conclusions: Our findings suggest a need for web-based activities for countering boredom and feelings of isolation. The main factors that influence the use, efficacy, and sustainability of online activities are access, motivational and need-fulfilling factors, and whether the activities are sufficiently tailored to individuals' preferences and abilities. Challenges in substituting in-person services are promoting social relationships that are currently not sufficiently incorporated into most web-based programs, accommodating a wider range of topics, and increasing the accessibility of current programs to older adults, especially those who are homebound, both during and after the COVID-19 pandemic.

(J Med Internet Res 2021;23(1):e25848) doi:10.2196/25848

KEYWORDS

web-based venues for older adults; social engagement; activities; boredom; technology barriers for seniors; COVID-19; pandemic; senior; elderly; older adult; online activity; engagement; activity; loneliness; isolation; effectiveness

Introduction

Background

The COVID-19 pandemic has harsh implications for the quality of life of older adults. Stay-at-home orders, closure of senior centers, and restrictions on visits by friends and relatives have increased social isolation and loneliness in this population [1-3]. Some of the stressors experienced by older adults as a result of the COVID-19 pandemic are related to being confined to home, concern for the health and safety of family and friends, and boredom [4]. The latter has also been reported as an impact of quarantine restrictions [5]. Homebound older adults, who are unable to leave the house due to illness or other impairments, may be at higher risk of feeling lonely during the pandemic [6] and at other times [7,8].

Social isolation and loneliness increase older people's risk of anxiety, depression [9], mortality [10], and dementia [11]. Activity and social engagement are important for psychological well-being [12], training of memory and executive function [13], greater happiness, and reduced mortality [14] of older adults.

Web-based technologies have been proposed as a way to address these issues while protecting older adults from COVID-19 infection [2,3]. These technologies are also cost-effective [15] and may benefit older adults in particular [16] because their social networks tend to be geographically less proximal [17].

Two categories of barriers have been identified concerning internet use among older adults: First, personal characteristics such as cognitive and physical impairments may limit their ability to use conventionally designed computer equipment [18,19]. Socioeconomic, educational, and cultural backgrounds influence older adults' ability to access computers as well as the extent of their pre-existing knowledge of technology and their experience using it [20,21]. Living arrangements are also an important factor [22], as people who live alone are less likely to use the internet [20], and complete beginners and frail older adults require extensive support and assistance [23].

Second, attitudes of older adults toward the internet and technology, including computer anxiety [24] and concern about data security and privacy, can present obstacles [25,26]. Older adults may find themselves excluded from the "digital world" due to lack of knowledge and skill in the use of modern technologies, which stems from a lack of assistance and training as well as from technology designs that fail to consider their needs, knowledge, and background [27-29].

Studies investigating the impact of pilot social web-based interventions reported positive effects on loneliness [30,31]. Information and communication technologies were found to reduce depression [32], and smart technologies increased self-efficacy, empowerment, and confidence in using technology [33].

However, most of these studies are qualitative, with small sample sizes; thus, they may be less conclusive. In addition, the positive impact of these interventions on social support and connectedness has been found to be short in duration [34]. Although the frequency of internet use was associated with reduced loneliness, it did not impact perceptions of social isolation [35]. The type of web-based activity influences impact, as only social activities (eg, connecting with family and friends) were associated with decreased loneliness [36] and enhanced life satisfaction [37], whereas internet use for informational purposes or instrumental functions (eg, banking) were not. Recreational activities were the only type of activity with a significant correlation with older adults' well-being after controlling for background variables [38].

Research on web-based group activities for older adults is scarce. We found only 2 studies that examined the impact of web-based physical exercise activities in different settings, comparing individual and group training [39,40]. Baez et al [39] reported high usability of a group intervention but did not find a decrease in loneliness attributable to the group intervention itself. Importantly, the web-based group exercises in these two studies [39,40] included avatars in a virtual gym instead of live video communication, which may have impacted the perceived quality of social contact and hence loneliness.

Aims of the Study

There is a need to further investigate the potential of web-based activities to enhance the well-being of older adults, particularly during the COVID-19 pandemic, where personal contact and interaction are severely restricted. Our specific research foci are (1) reasons older persons use or do not use web-based activities; (2) the effectiveness of various web-based activities in achieving user satisfaction and the use of alternative web-based activities; and (3) older persons' activity preferences and their willingness to pay for web-based activities.

Methods

The Healthy Aging Web-Based Activity Program

An opportunity to examine these issues arose when we heard that after Healthy Aging, Ltd (*Beseva Bria*), a for-profit organization providing rehabilitation services to older persons, was forced to close its doors due to COVID-19 regulations, they started providing activities via the internet platform Zoom (Zoom Video Communications), which facilitates video group meetings. Web-based activities were initially offered free of charge when staff volunteered to provide older persons with activities and companionship after senior centers closed during the COVID-19 crisis. Thereafter, these Zoom meetings were offered for a small subscription fee.

Healthy Aging's Zoom activities took place 5 days per week, with 3 activities per day, each lasting 30 minutes, between 10 AM and 11:30 AM. The first activity was always a type of exercise that could be performed while seated.

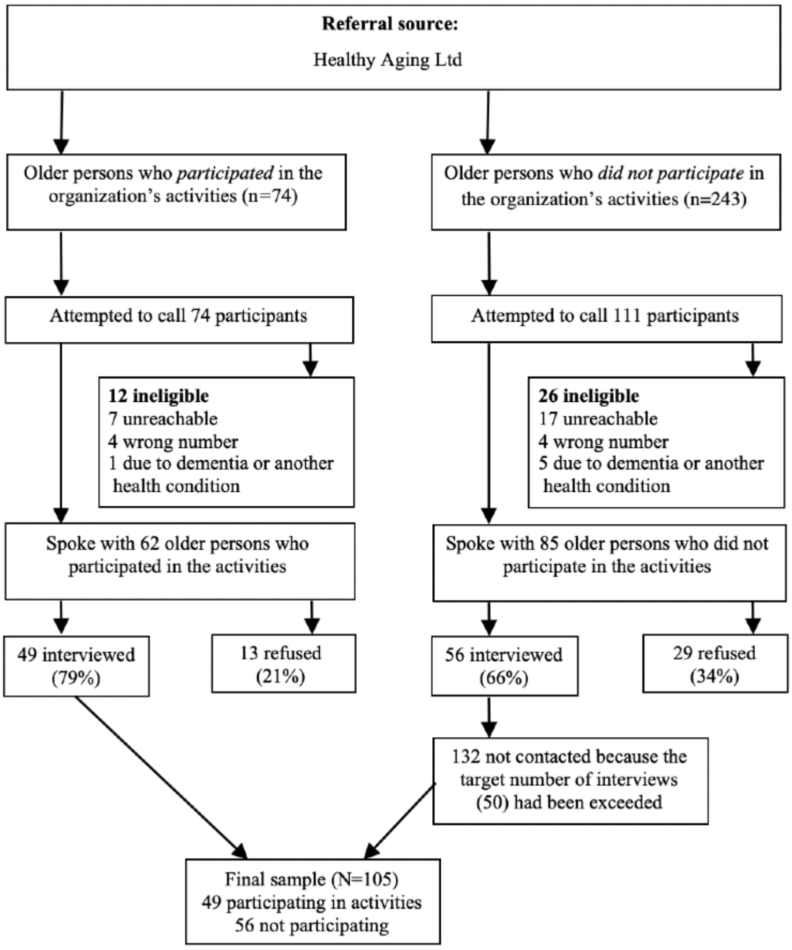
The other activities varied, including mindfulness, musical tai-chi, self-help, and lectures about topics such as world travel, history, health, and mental health. When deemed appropriate by Healthy Aging, some participants were invited to present lectures. Starting at 9:30 AM, some participants engaged in an informal chat via the same Zoom meeting. In the lectures and mindfulness activities, participants were encouraged to participate in discussions. The vast majority of the participants did not know each other, nor did they know the organization prior to connecting to the web-based program.

Recruitment

After Healthy Aging agreed, we embarked on a pilot study to examine the utility of Zoom activities for the older population served by the organization. Ethical approval was obtained from the Institutional Review Board of Tel Aviv University.

Healthy Aging sent letters informing prospective interviewees of this study and offering them the opportunity to opt out of being contacted for it. Letters were sent to all persons who participated in the organization's web-based activities or were offered the opportunity to participate. Healthy Aging then provided us with their contact information. The recruitment process is described in Figure 1. Although the percentage of people who declined to be interviewed was higher among people who did not participate in the activities (34%) than among activity participants (21%), this difference was not statistically significant (*P*=.08). We interviewed 49 participants and 56 nonparticipants.

Figure 1. Participant recruitment process for the study.



Assessments

Separate questionnaires were developed for participants and nonparticipants based on our research foci (Multimedia Appendix 1 and Multimedia Appendix 2, respectively). Interviews for both groups were conducted via telephone and lasted for approximately 20 minutes. Both interviews started with an explanation of the interview and a request for verbal informed consent. Both questionnaires included queries about the respondent's demographic background. When someone other than the older person was the interviewee, questions were asked about the reason for this and the relationship of the interviewee to the older person.

The remainder of the questionnaire for activity participants included 37 closed- and open-ended questions, such as “How often have you participated in the activities?” rated on a scale of less than once a week, once or twice a week, three or four days a week, and every day. Comments offered in response to open-ended questions were transcribed. The topics covered were ease of using technology, frequency of participation, content of the activities, reasons for participation and their importance, satisfaction with particular activities, participation in alternative activities, desire for promotion of social relationships through the Zoom activities, and willingness to pay for web-based activities.

The questionnaire for respondents who did not participate in the activities included 25 similar questions, including their reasons for nonparticipation.

Analytic Approach

Statistical analysis involved descriptive statistics using SPSS (IBM Corporation). The two groups were compared via *t* test for ordinal and interval data and via chi-square test for nominal-level data. When answers to 2 questions overlapped, such as “Why do you not participate in Healthy Aging’s online activities?” and “What changes or improvements in the online format would have motivated you to participate in particular activities?” the responses were combined. Open-ended responses were coded using an emergent coding strategy [41] whereby two research staff members read, coded, and categorized the responses independently and then revised the codes through discussion until agreement was reached. Their findings were reviewed by another staff member, and the main themes were ultimately agreed upon by all researchers. Open-ended responses are presented as quotations where they inform the quantitative findings.

Results

Demographics

As shown in Table 1, both groups had a mean age of approximately 74 years (participants, mean 74.3 years, SD 6.6; nonparticipants, mean 74.8 years, SD 8.7) and approximately 15 years of education (participants, mean 15.3 years, SD 4.2; nonparticipants, mean 15.2 years, SD 3.6). Over 80% of the respondents (89/105, 84.8%) were female. Over 90% (92/97, 95%) lived in their homes, and 60% (57/95) lived with a spouse. Demographic differences between those who participated in the web-based activities and those who did not were generally not significant. There was a trend for a larger percentage of nonparticipants (8/50, 16%) to require help in walking, compared to 2/47 participants (4%; *P*=.06) and in requiring help reaching places outside of walking distance (10/51 nonparticipants, 20%, vs 3/47 participants, 6%; *P*=.054). In 10/56 interviews (18%) with nonparticipants, someone other than the older person completed the interview, compared to 2% (1/49) in the case of participants (*P*=.008). In 13% (7/46) of interviews with nonparticipants who answered for themselves, the interviewer thought the respondent had some cognitive difficulties, compared to 2% (1/49) among participants (*P*=.04).

Table 1. Background variables of the study participants (N=105).

Variable	Participating in activities (n=49)	Not participating in activities (n=56)	Total sample (N=105)	Difference between groups (*P* value)
Age (years), mean (SD)	74.3 (6.6)	74.8 (8.7)	74.6 (7.8)	.77
Education (years), mean (SD)	15.3 (4.2)	15.2 (3.6)	15.3 (3.9)	.91
Female gender, n (%)	43/49 (87.8)	46/56 (82.1)	88/105 (84.8)	.42
Born in Israel, n (%)	26/47 (55.3)	32/52 (61.5)	58/99 (58.6)	.53
Married (%)	27/46 (58.7)	28/51(54.9)	55/97 (56.7)	.71
Residence (%)	43/45 (95.6)	49/52 (94.2)	92/97 (94.8)	.77
Living situation, n (%)				.55
With spouse	27/45 (60.0)	30/50 (60.0)	57/95 (60.0)	
Alone	15/45 (33.3)	16/50 (32.0)	31/95 (32.6)	
With caregiver	3/45 (6.7)	2/50 (4.0)	5/95 (5.3)	
With other family member	0/50 (0.0)	2/50 (4.0)	2/95 (2.1)	
Can walk without help, n (%)	45/47 (95.7)	42/50 (84.0)	87/97 (89.7)	.06[a]
Can reach places farther than walking distance without help, n (%)	44/47 (93.6)	41/ 51(80.4)	85/98 (86.7)	.054[b]
Working, n (%)	9/46 (19.6)	16/52 (30.8)	25/98 (25.5)	.20
Older person answered questions themselves, n (%)	48/49 (98.0)	46/56 (82.1)	94/105 (89.5)	.008[c]
Impression of interviewer: cognitive impairment of person interviewed “Not at all,” n (%)	48/49 (98.0)	49/56 (87.5)	97/ 105 (92.4)	.04[d]
Impression of interviewer: accurate information given, n (%)	49/49 (100.0)	53/56 (94.6)	102/105 (97.1)	.10

[a] χ^2_1=3.6.

[b] χ^2_1=3.7.

[c] χ^2_1=7.0.

[d] χ^2_1=4.1.

Feasibility: Extent of and Reasons for Participation and Nonparticipation

Of the 49 participants who participated in the activities, 27 (55%) reported participating in the activities every day, with 11 (22%) reporting participation 3 or 4 days per week; 37 (76%) had participated in the activities for over 20 days at the time of the interview.

Most of the 49 participants (34, 69%) were able to access Zoom on their own; however, 15 (31%) had difficulties, mostly in activating Zoom and starting the activities. These participants were helped by family members (9/49, 18%); by staff from Healthy Aging (4/49, 8%); by paid caregivers (1/49, 2%); or by a hired technician (1/49, 2%).

Respondents who participated in the activities rated enjoying the content of the activity as the most important motivator (mean score 4.4 on a scale of 1-5); as one participant commented, “This way you are exposed to interesting lectures and new people” [#162, age 74 years, female]. This was followed by maintaining a routine (mean score 3.6), as in “[It helps me] get up on time. …It provides me with a framework and routine” [#131, age 77 years, female]; enjoying the group and the presence of others (mean score 3.1); relief from loneliness (mean score 2.6), such as “Company during a time of loneliness” [#129, age 73 years, female]; and being motivated by family members or friends (mean score 1.8). One participant commented that the web-based activities helped her avoid loneliness and depression (#133, age 87 years, female).

Similar to the above ratings, the most common reasons for engaging in the web-based activities reported by participants were interest in the activities and relief from boredom for 30/49 (61%) (Table 2), followed by opportunity to exercise, access to activities from home, and maintaining a daily routine. Only 8/49 participants (16%) mentioned social activity or relief from loneliness as a reason for participating.

Table 2. Study participants' reasons for participating (n=49) and not participating (n=56), n=105.[a]

Category and reason	Participants, n (%)
Reasons for participating (n=49), 82 total responses	
Motivation	
Interest in content/relieving boredom	30 (61)
Providing exercise	15 (31)
Maintaining a daily routine	11 (22)
Social interest/relief of loneliness	8 (16)
Activities appropriate for older persons	2 (4)
Access	
Activities accessible from home	14 (29)
Convenient hours	2 (4)
Reasons for not participating (n=56), 102 total responses	
Access	
Never heard of Healthy Aging	24 (42)
Technical issues[b]	13 (23)
Inconvenient time or duration of the activities[c]	12 (21)
Not willing to pay	3 (5)
Motivation	
Lack of interest in the content of the activities	18 (32)
Participation in other activities/organizations	9 (16)
The activities are designed for older persons ("I'm too young for these")	5 (9)
Abilities	
Problems due to cognitive impairment[d]	7 (13)
Hearing/vision problems[d]	4 (7)
Concerns	
Reluctance to take part in group activities	3 (5)
Reluctance to participate in activities with a camera	1 (2)
The organization seems too commercial	1 (2)

[a] Participants could provide multiple answers.

[b] Out of 13 participants, 8 (62%) did not know how to operate a computer/phone, 3 (23%) indicated they did not relate to technology and Zoom, and 2 (15%) had no computer.

[c] Out of 12 participants, 8 (67%) reported participating in other activities, 1 (8%) reported working, and 1 (8%) reported having no time due to caregiving for a spouse with dementia.

[d] Two persons reported cognitive reasons *and* hearing/vision problems.

Nonparticipants gave an average of 1.8 reasons for not participating (range 1-4, SD 0.9). The most common reasons were related to lack of access and awareness, such as not having heard of the activities sponsored by Healthy Aging (24/56, 43%) and technical issues (13/56, 23%), such as not owning a computer and inability or lack of knowledge of how to connect to Zoom. Motivational factors, such as lack of interest in the content of the offered activities, also played a role (Table 2), as did participants' involvement in other activities (eg, "The activities offered are not as good as other activities I have and do" [#214, age 61 years, female]). Other reasons reported less often were cognitive and sensory problems and the perception that the activity was appropriate for an older group (Table 2). Respondents who felt the activity was appropriate for older persons expressed a preference for activities such as home repairs, yoga, and belly dancing. The average age of these respondents was 72.8 years (range 66-85 years), compared to 74.6 years for the full sample.

Efficacy and Avenues for Upgrade: Participants' Satisfaction With Web-Based Activities and Their Ideas for Improvement

As shown in Table 3, exercise was rated as the most satisfying activity (4.4 on a 5-point scale), and these activities were attended by most participants: "There is great diversity, every day, a different part of the body" [#110, age 75 years, female]. Mindfulness received the lowest ratings, with an average score of 3, denoting moderate satisfaction; one participant stated, "I do not relate to it" [#112, age 77 years, female]. The ordinal order of the levels of satisfaction for the different activities is roughly reflected in the attendance levels (Table 3), with the highest attendance at exercise and professional lecturer activities. The lowest attendance was reported for lectures provided by group members.

Out of the 49 activity participants, 28 (57%) expressed interest in continued participation in exercise activities after the COVID-19 pandemic was over, followed by interest in lectures, while other activities drew less interest. These reflect the same order of satisfaction with the activities (Table 3).

Table 3. Reported levels of satisfaction and desire to continue each type of activity offered by Healthy Aging after the COVID-19 pandemic (n=49).

Type of activity	Reported satisfaction levels per type of activity,[a] mean (SD)	Number attending, n	Would like to continue the activity after the pandemic, n (%)
Physical exercise	4.4 (0.7)	44	28 (57)
Lecture[b] (professional lecturer)	4.3 (1.0)	37	20 (41)
Lecture (lecturer from the participants' group)	3.8 (1.0)	17	13 (27)
Self-care exercises (eg, head massage)	3.8 (1.4)	26	10 (20)
"Travel from the couch" lecture by a tour guide	3.7 (1.1)	25	14 (29)
Mindfulness	3.0 (1.5)	24	7 (14)

[a] 1, not at all; 2, a little; 3, moderately satisfied; 4, satisfied; 5, very satisfied.

[b] Lectures included diverse topics, such as current art, memory, nutrition, the Bible, sexuality in old age, and history.

Participants indicated several ways through which their interest in participating in the activities could be increased, such as enhancing social contacts, enriching the content of activities, and improving technical and scheduling features. In terms of encouraging social contact, 42% (20/48; one participant did not answer this question) thought that social contact should be a goal of activities. In terms of the content of the activities, 33% of participants (16/49) viewed it as insufficiently interesting; in contrast, 2% (1/49) described the content as too complicated. Moreover, 2/49 participants (4%) considered the delivery to be too fast, and the same number requested more pictures and music in the presentations (2/49, 4%). Requests for additional types of content were especially common (13/49, 27%), as presented in Table 4, which includes requested content expressed by participants and by nonparticipants. The most popular specific topics were exercise, such as "I love Pilates" [#296, age 72 years, female]; culture, such as "lectures in museums" [#121, age 85 years, male]; music, such as "series of lectures about jazz" [#108, age 78 years, female]; art, such as "activities in the field of painting" [#224, age 86 years, male], and travel, such as "lectures on trips around the world" [#225, age 85 years, male]. However, this is not a perfect categorization, because many participants requested very specific activities, such as "play Remi," "Nordic walking," or "psychology of the brain." Finally, 3 out of 49 participants (6%) mentioned problems with their computer, Zoom, or sound quality, and some requested simpler presentation formats, such as television. Out of the 49 activity participants, 10 (20%) sought programming at different times and during more time slots or recording of activities for later viewing.

Table 4. Preferred areas of interest of participants (n=49) and nonparticipants (n=56) in web-based activities (n=105).

Total (N=105), n (%)	Nonparticipants (n=56), n (%)	Participants (n=49), n (%)	Topic
33 (31.4)	13 (23.2)	20 (40.8)	Lectures—general
29 (27.6)	15 (26.8)	14 (28.6)	Exercise
25 (23.8)	10 (17.9)	15 (30.6)	Art/culture/music
12 (11.4)	6 (10.7)	6 (12.2)	Travel
6 (5.7)	2 (3.6)	4 (8.2)	History/philosophy
6 (5.7)	3 (5.4)	3 (6.1)	Coaching/body-mind
19 (18.1)	6 (10.7)	13 (26.5)	Other[a]

[a] Includes games/mind games/bridge (4 participants and 1 nonparticipant), literature (2 participants and 1 nonparticipant), science/technology (3 participants and 0 nonparticipants), education/social sciences (1 participant and 2 nonparticipants), religion (2 participants and 0 nonparticipants), food/cooking (0 participants and 2 nonparticipants), language study (1 participant and 1 nonparticipant), current affairs (1 participant and 0 nonparticipants), business (0 participants and 1 nonparticipant), gardening (1 participant and 0 nonparticipants), and wills (1 participant and 0 nonparticipants).

Sustainability: Competing Activities and Willingness to Pay for Activities

Web-based activity participants were significantly more likely to report partaking in alternative activities (38/48, 79%) than nonparticipants (34/56, 61%; χ^2_1=4.1; P=.04). Of those participating in Healthy Aging's web-based activities, 10/38 (21%) reported not participating in other activities. Of those not involved with Healthy Aging web-based activities, 22/56 (39%) reported not participating in any activities. Of these 22 individuals, 10 (45%) reported either cognitive, sensory, or technological problems. Alternative activities (Table 5) were accessed via television, such as a show with exercise instruction, or YouTube, Zoom, and websites of different organizations. Some participants reported that the activity was a continuation of a class they had taken in person prior to the COVID-19 pandemic. Sponsors of alternative activities varied, such as universities (eg, "Lectures on a film, and then [watching] the film" [#110, age 75 years, female]), museums (eg, "I have a subscription… and I listen to their lectures" [#120, age 63 years, female]), municipalities, synagogues (eg, "Prayer in the Synagogue via Zoom" [#103, age 79 years, female]), not-for-profit and for-profit organizations, and private individuals.

Table 5. Characteristics of the alternative activities engaged in by the study participants (n=73).

Activity or provider	Participants (n=39), n (%)	Nonparticipants (n=34), n (%)	Total (n=73), n (%)
Platform			
Zoom	20 (51)	8 (24)	28 (38)
Television	1 (3)	2 (6)	3 (4)
YouTube	1 (3)	4 (12)	5 (7)
Organization website	7 (18)	4 (12)	11 (15)
Provider of activity			
University/museum	9 (23)	8 (24)	17 (23)
Municipality	13 (33)	3 (9)	16 (22)
Synagogue	1 (3)	0 (0)	1 (1)
Not-for-profit organization	3 (8)	3 (9)	6 (8)
Private person/organization	26 (67)	22 (65)	48 (66)

Of 47 respondents participating in web-based activities, 60% (n=28) said they were willing to pay for them, and 26% (n=12) were unwilling; meanwhile, 6 participants (13%) said "maybe," and 1 (2%) said "I don't know." Among the interviewees who participated in the beginning of the program, 7 said they had stopped participating after Healthy Aging introduced a small subscription fee, while 14 reported that they were paying for participation at the time of the interview. Those who paid cited fairness and wanting or needing the service:

> *It's worth it to me, it's fair, and you get a lot for it* [#103, age 79 years, female]
>
> *I have no choice. If not for the activities, I'll be lost… So as not to be left alone [I choose to continue]* [#133, age 87 years, female]

Reasons for not wanting to pay varied, including the availability of alternatives (eg, "So far, I didn't feel the need to pay because there are a lot of options" [#138, age 68 years, female]); an ideological assertion that such services should be free at the time of a pandemic (eg, "Healthy Aging began [the program] nicely but then asked for money. I think it should continue for free until the end of COVID" [#121, age 85 years, male]); and "double billing":

> *I used to participate in the exercise of Healthy Aging, but stopped when they required payment. I am not willing to pay because I am still paying a lot to my gym, despite not being able to go there.* [#116, age 79 years, female]

Out of the 56 nonparticipants, 20 (36%) indicated willingness to pay for Zoom activities; however, 12 (26%) responded in the negative. Although these percentages reflect less readiness to pay for such activities than reported by the activity participants, the difference is not statistically significant (P=.33). Those who were willing to pay tended to have more years of education (mean 15.8 years, SD 3.5) than those who said they would not or did not know if they would pay (mean 14.2 years, SD 3.6; P=.03).

Discussion

Principal Findings

Participants in Healthy Aging's web-based activities reported very high levels of satisfaction with the exercise and lecture programs. Over 50% of the participants (28/49, 57.1%) wished to continue with the exercise sessions after the end of the COVID-19 pandemic, and over 40% (20/49, 40.8%) wished to continue with the lectures, citing the benefits of maintaining physical vitality, participating in interesting activities, and maintaining a daily routine. These results suggest that web-based activities are a viable substitute for the pre–COVID-19 activities of older adults. This finding aligns with Whitehead and Torossian's [4] report that digital social contact was among the most commonly reported sources of joy or comfort for older adults during the COVID-19 pandemic. However, the results of our sample suggested insufficient provision of social contact and preventing loneliness—a benefit mentioned by only 16%

of the participants (8/49). The lack of social cohesion may have also been reflected by the low attendance when lectures were presented by other participants. This finding supports previous studies that examined web-based exercise group activities and found no impact on loneliness [39,40,42] and research results indicating that older adults generally prefer face-to-face interactions [43,44]. The extent to and the conditions under which web-based activities can prevent loneliness should be explored in future research. However, the lack of social interaction in web-based activities is not unlike that at senior centers, which provide live lectures, concerts, or exercise classes; many participants arrive to these activities alone and leave without having had significant social interaction [45]. Strategies to promote social interaction both on the web and in person should be examined in future research.

Conceptual Framework

Based on our findings, we have developed a conceptual framework that captures factors affecting the use, efficacy, and sustainability of web-based activities, as summarized in Table 6.

Table 6. Conceptual model: Factors affecting the use, efficacy, and sustainability of web-based activities.

Category	Factors
Factors affecting use	
Access	• Knowledge about program availability • Sensory abilities (natural or modified) • Technology access (personal or assisted) • Sufficient funds to pay for subscription (real or perceived) • Convenient hours and duration
Concerns	• Group activities • Privacy: use of camera
Factors affecting use, efficacy, and sustainability	
Fulfilling needs/motivational	• Relieving boredom • Providing physical activity • Providing social interaction, relief of loneliness
Tailoring activities to individuals’ preferences and abilities	• Topic fit (interest) • Level fit (difficulty match) • Physical abilities • Cognitive abilities • Educational level • Presentation fit (speed, timing, etc)

The main factors influencing the use of web-based resources were access and awareness. A sizable proportion of nonparticipants denied having heard of Healthy Aging’s web-based activities, despite having been notified about them by Healthy Aging. Further study should explore how to effectively promote and advertise web-based activities for older persons to ensure knowledge and comprehension of the activities by prospective users.

Another component of access was the technological challenges web-based activities pose for older persons, as reported by 31% (15/49) of the web-based activity participants—who nevertheless managed to participate—and 23% (13/56) of the nonparticipants, for whom technology was a complete barrier to participation. While this barrier might be overcome through the help of family members, caregivers, or staff from organizations, lack of such support remains a major barrier to internet and computer use by older persons [23]. Simplifying the process of use and adapting it to cognitive and sensory limitations may also aid in overcoming technological barriers. Indeed, sensory abilities, natural or modified, also affected participants’ access to web-based activities, as reported by 7% of nonparticipants (4/56). The extent to which compensation for these limitations can help disabled persons benefit from web-based activities needs to be examined.

Half of the nonparticipants (28/56, 50%) and 60% (28/47) of the participants said they would be willing to pay for web-based programming. Others were not inclined to pay for a range of reasons, including the availability of cost-free programs. A portion of this population may not be able to afford subscription fees.

Certain concerns emerged as factors that affected use. Some persons in our sample were reluctant to participate in activities involving a camera or a group. Such privacy concerns were previously reported as reasons for older adults’ hesitation to use the internet [25,26].

We identified physical activity, social interaction, and relief from boredom and loneliness as central motivational and needs-fulfilling factors that affected the use, efficacy, and sustainability of web-based activities. The preference for group activities, the importance of social interaction [40], and the opportunity to exercise were identified as motivators in previous research on in-person group exercises for older adults [46].

The final factor that we identified was tailoring activities to individuals’ preferences and abilities. The activities provided by Healthy Aging were an extension of its mission as a rehabilitation facility. Accordingly, programs focused on activities such as seated physical exercise, head massage, and stimulating lectures. However, for some respondents, these types of programs were considered to be relevant only to an older, frailer population.

Both participants and nonparticipants were interested in a wide range of topics, and many were very particular about favored topics, suggesting the commercial feasibility of developing wide-ranging lecture series; 50% of our sample were willing to pay for web-based activities. The suitability of the activity topics and content in our sample emerged as an important factor not previously identified in the literature.

The match between activity participation and participants' cognitive and physical abilities as well as educational background proved to be critical. Cognitive or sensory problems interfered with the ability of some respondents to benefit from the web-based activities, consistent with prior literature [18,19]. Cognitive issues appeared to be more prevalent for nonparticipants, for whom both the percentage of interviews via third persons and the percentage of interviewees for whom interviewers suspected cognitive difficulty were significantly greater than in the case of the web-based activity participants.

Limitations

This study is limited by its relatively small sample that is confined to one country and by its focus on a single rehabilitation organization that developed web-based activities for older persons during the COVID-19 pandemic. However, this narrow focus afforded us important insights into the experience of those who participated and those who declined as well as into the future potential of web-based programs for older persons.

Conclusions

The closing of social clubs, senior centers, libraries, and gymnasiums due to COVID-19 critically affected the living experiences of older persons, who were instructed to limit social contact, stay at home, and decline visitors. Television and web-based activities were the main venues left to help older persons remain engaged and somewhat active. Our findings suggest that this alternative provides valuable substitute activities to a large portion of older persons who were required to limit outdoor activities due to the COVID-19 pandemic.

This study provides a preliminary investigation of the feasibility, efficacy, and sustainability of web-based activities among older persons. Each of these aspects of the study deserves a study of its own. Feasibility issues pertain to how to improve publicity and access to web-based programming. Efficacy relates to how to improve the match between the activities and the participants' needs and wishes, as well as how to facilitate social interaction and to decrease loneliness. Finally, sustainability raises issues regarding funding, including participant contributions, to develop and present high quality activities that appeal to older persons.

In terms of enhancing access to web-based activities, training and support are needed for some older persons to access web-based programs [47,48]. In addition, it has been argued that improving older adults' access to technology involves reducing the costs of computers, smartphones, and other equipment as well as adjusting their design to be more user-friendly to older adults [49]. As Seifert et al [49] aptly stated, older adults are impacted by "double exclusion"—social and digital—during the COVID-19 pandemic. These issues need to be explored in future research into the most effective means for increasing access to older populations with different types of limitations.

The results of this study have implications beyond the COVID-19 pandemic as a modality to address the needs of homebound older persons. The main challenges in substituting in-person services include the challenges of promoting social relationships within web-based platforms, accommodating a wider range of activities and contents on the web, and making current programs accessible to an underserved population that needs them, both through better marketing and via improving access to technology.

Acknowledgments

We would like to thank the Healthy Aging staff and study participants for their cooperation in facilitating this study. This work was supported by the Minerva Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire for older adults participating in the web-based activities.
[DOCX File , 33 KB - jmir_v23i1e25848_app1.docx]

Multimedia Appendix 2
Questionnaire for older adults who did not participate in the web-based activities.
[DOCX File , 28 KB - jmir_v23i1e25848_app2.docx]

References

1. Berg-Weger M, Morley JE. Editorial: Loneliness and Social Isolation in Older Adults during the COVID-19 Pandemic: Implications for Gerontological Social Work. J Nutr Health Aging 2020;24(5):456-458 [FREE Full text] [doi: 10.1007/s12603-020-1366-8] [Medline: 32346678]
2. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health 2020 May;5(5):e256 [FREE Full text] [doi: 10.1016/S2468-2667(20)30061-X] [Medline: 32199471]
3. Brooke J, Jackson D. Older people and COVID-19: Isolation, risk and ageism. J Clin Nurs 2020 Jul;29(13-14):2044-2046. [doi: 10.1111/jocn.15274] [Medline: 32239784]
4. Whitehead BR, Torossian E. Older Adults' Experience of the COVID-19 Pandemic:A Mixed-Methods Analysis of Stresses and Joys. Gerontologist 2020 Sep 04:gnaa126 [FREE Full text] [doi: 10.1093/geront/gnaa126] [Medline: 32886764]
5. Goodman-Casanova JM, Dura-Perez E, Guzman-Parra J, Cuesta-Vargas A, Mayoral-Cleries F. Telehealth Home Support During COVID-19 Confinement for Community-Dwelling Older Adults With Mild Cognitive Impairment or Mild Dementia: Survey Study. J Med Internet Res 2020 May 22;22(5):e19434 [FREE Full text] [doi: 10.2196/19434] [Medline: 32401215]
6. Donovan NJ. Timely Insights Into the Treatment of Social Disconnection in Lonely, Homebound Older Adults. Am J Geriatr Psychiatry 2020 Jul;28(7):709-711 [FREE Full text] [doi: 10.1016/j.jagp.2020.04.002] [Medline: 32303405]
7. Cohen-Mansfield J, Shmotkin D, Hazan H. Homebound older persons: prevalence, characteristics, and longitudinal predictors. Arch Gerontol Geriatr 2012 Jan;54(1):55-60 [FREE Full text] [doi: 10.1016/j.archger.2011.02.016] [Medline: 21420181]
8. Cohen-Mansfield J, Shmotkin D, Hazan H. The effect of homebound status on older persons. J Am Geriatr Soc 2010 Dec;58(12):2358-2362 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03172.x] [Medline: 21087220]
9. Santini ZI, Jose PE, York Cornwell E, Koyanagi A, Nielsen L, Hinrichsen C, et al. Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. Lancet Public Health 2020 Jan;5(1):e62-e70 [FREE Full text] [doi: 10.1016/S2468-2667(19)30230-0] [Medline: 31910981]
10. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci USA 2013 Apr 09;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]
11. Sundström A, Adolfsson AN, Nordin M, Adolfsson R. Loneliness Increases the Risk of All-Cause Dementia and Alzheimer's Disease. J Gerontol B Psychol Sci Soc Sci 2020 Apr 16;75(5):919-926 [FREE Full text] [doi: 10.1093/geronb/gbz139] [Medline: 31676909]
12. Park NS. The Relationship of Social Engagement to Psychological Well-Being of Older Adults in Assisted Living Facilities. J Appl Gerontol 2009 Feb 03;28(4):461-481. [doi: 10.1177/0733464808328606]
13. Carlson MC, Saczynski JS, Rebok GW, Seeman T, Glass TA, McGill S, et al. Exploring the effects of an "everyday" activity program on executive function and memory in older adults: Experience Corps. Gerontologist 2008 Dec;48(6):793-801. [doi: 10.1093/geront/48.6.793] [Medline: 19139252]
14. Menec VH. The relation between everyday activities and successful aging: a 6-year longitudinal study. J Gerontol B Psychol Sci Soc Sci 2003 Mar;58(2):S74-S82. [doi: 10.1093/geronb/58.2.s74] [Medline: 12646596]
15. Aalbers T, Baars MAE, Rikkert MGMO. Characteristics of effective Internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
16. Hülür G, Macdonald B. Rethinking social relationships in old age: Digitalization and the social lives of older adults. Am Psychol 2020;75(4):554-566. [doi: 10.1037/amp0000604] [Medline: 32378949]
17. Ajrouch KJ, Antonucci TC, Janevic MR. Social networks among blacks and whites: the interaction between race and age. J Gerontol B Psychol Sci Soc Sci 2001 Mar;56(2):S112-S118. [doi: 10.1093/geronb/56.2.s112] [Medline: 11245365]
18. Good A, Stokes S, Jerrams-Smith J. Elderly, novice users and health information web sites: issues of accessibility and usability. J Healthc Inf Manag 2007;21(3):72-79. [Medline: 19195297]
19. Namazi KH, McClintic M. Computer use among elderly persons in long-term care facilities. Educ Gerontol 2003 Jun;29(6):535-550. [doi: 10.1080/713844391]
20. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
21. Anderson M, Perrin A, Jiang J, Kumar M. 10% of Americans don’t use the internet. Who are they? Pew Research Center. 2019 Apr 22. URL: https://www.pewresearch.org/fact-tank/2019/04/22/some-americans-dont-use-the-internet-who-are-they/ [accessed 2020-10-27]
22. Leukel J, Schehl B, Sugumaran V. To Do or Not to Do: How Socio-demographic Characteristics of Older Adults Are Associated with Online Activities. In: Gao Q, Zhou J, editors. Human Aspects of IT for the Aged Population. Technology and Society. Cham, Switzerland: Springer; Jul 10, 2020.
23. Dickinson A, Gregor P. Computer use has no demonstrated impact on the well-being of older adults. Int J Hum Comput Stud 2006 Aug;64(8):744-753. [doi: 10.1016/j.ijhcs.2006.03.001]

24. Jung Y, Peng W, Moran M, Jin SA, McLaughlin M, Cody M, et al. Low-Income Minority Seniors' Enrollment in a Cybercafé: Psychological Barriers to Crossing the Digital Divide. Educ Gerontol 2010 Feb 11;36(3):193-212. [doi: 10.1080/03601270903183313]
25. Hussain D, Ross P, Bednar P. The Perception of the Benefits and Drawbacks of Internet Usage by the Elderly People. In: Rossignoli C, Virili F, Za S, editors. Digital Technology and Organizational Change. Cham, Switzerland: Springer; 2018:199-212.
26. Bixter MT, Blocker KA, Mitzner TL, Prakash A, Rogers WA. Understanding the use and non-use of social communication technologies by older adults: A qualitative test and extension of the UTAUT model. Gerontechnology 2019 Jun;18(2):70-88 [FREE Full text] [doi: 10.4017/gt.2019.18.2.002.00] [Medline: 31754352]
27. Seifert A, Reinwand DA, Schlomann A. Designing and Using Digital Mental Health Interventions for Older Adults: Being Aware of Digital Inequality. Front Psychiatry 2019 Aug 9;10:568 [FREE Full text] [doi: 10.3389/fpsyt.2019.00568] [Medline: 31447716]
28. Prensky M. Digital Natives, Digital Immigrants Part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
29. Schulz R, Wahl H, Matthews JT, De Vito Dabbs A, Beach SR, Czaja SJ. Advancing the Aging and Technology Agenda in Gerontology. Gerontologist 2015 Oct;55(5):724-734 [FREE Full text] [doi: 10.1093/geront/gnu071] [Medline: 25165042]
30. Larsson E, Padyab M, Larsson-Lund M, Nilsson I. Effects of a social internet-based intervention programme for older adults: An explorative randomised crossover study. British Journal of Occupational Therapy 2016 Jul 19;79(10):629-636. [doi: 10.1177/0308022616641701]
31. Ballantyne A, Trenwith L, Zubrinich S, Corlis M. ‘I feel less lonely’: what older people say about participating in a social networking website. Quality Ageing Older Adults 2010 Sep 13;11(3):25-35. [doi: 10.5042/qiaoa.2010.0526]
32. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among older adults. Comput Hum Behav 2012 Mar;28(2):496-499. [doi: 10.1016/j.chb.2011.10.021]
33. Barbosa Neves B, Franz R, Judges R, Beermann C, Baecker R. Can Digital Technology Enhance Social Connectedness Among Older Adults? A Feasibility Study. J Appl Gerontol 2019 Jan;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]
34. Chen YR, Schulz PJ. The Effect of Information Communication Technology Interventions on Reducing Social Isolation in the Elderly: A Systematic Review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
35. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
36. Szabo A, Allen J, Stephens C, Alpass F. Longitudinal Analysis of the Relationship Between Purposes of Internet Use and Well-being Among Older Adults. Gerontologist 2019 Jan 09;59(1):58-68. [doi: 10.1093/geront/gny036] [Medline: 29688332]
37. Lam SSM, Jivraj S, Scholes S. Exploring the Relationship Between Internet Use and Mental Health Among Older Adults in England: Longitudinal Observational Study. J Med Internet Res 2020 Jul 28;22(7):e15683 [FREE Full text] [doi: 10.2196/15683] [Medline: 32718913]
38. Lifshitz R, Nimrod G, Bachner YG. Internet use and well-being in later life: a functional approach. Aging Ment Health 2018 Jan;22(1):85-91. [doi: 10.1080/13607863.2016.1232370] [Medline: 27657190]
39. Nikitina S, Didino D, Baez M, Casati F. Feasibility of Virtual Tablet-Based Group Exercise Among Older Adults in Siberia: Findings From Two Pilot Trials. JMIR Mhealth Uhealth 2018 Feb 27;6(2):e40 [FREE Full text] [doi: 10.2196/mhealth.7531] [Medline: 29487045]
40. Baez M, Khaghani Far I, Ibarra F, Ferron M, Didino D, Casati F. Effects of online group exercises for older adults on physical, psychological and social wellbeing: a randomized pilot trial. PeerJ 2017;5:e3150 [FREE Full text] [doi: 10.7717/peerj.3150] [Medline: 28392983]
41. Stemler S. An overview of content analysis. Pract Assess Res Evaluation 2000;7(1):17. [doi: 10.7275/z6fm-2e34]
42. van Boekel LC, Peek ST, Luijkx KG. Diversity in Older Adults' Use of the Internet: Identifying Subgroups Through Latent Class Analysis. J Med Internet Res 2017 May 24;19(5):e180 [FREE Full text] [doi: 10.2196/jmir.6853] [Medline: 28539302]
43. Yuan S, Hussain SA, Hales KD, Cotten SR. What do they like? Communication preferences and patterns of older adults in the United States: The role of technology. Educ Gerontol 2015 Oct 02;42(3):163-174. [doi: 10.1080/03601277.2015.1083392]
44. Andrews JA, Brown LJ, Hawley MS, Astell AJ. Older Adults' Perspectives on Using Digital Technology to Maintain Good Mental Health: Interactive Group Study. J Med Internet Res 2019 Feb 13;21(2):e11694 [FREE Full text] [doi: 10.2196/11694] [Medline: 30758292]
45. Cohen-Mansfield J, Hazan H, Lerman Y, Shalom V, Birkenfeld S, Cohen R. Efficacy of the I-SOCIAL intervention for loneliness in old age: Lessons from a randomized controlled trial. J Psychiatr Res 2018 Apr;99:69-75. [doi: 10.1016/j.jpsychires.2018.01.014] [Medline: 29407289]
46. de Groot GCL, Fagerström L. Older adults' motivating factors and barriers to exercise to prevent falls. Scand J Occup Ther 2011 Jun;18(2):153-160. [doi: 10.3109/11038128.2010.487113] [Medline: 20545467]

47. Betts LR, Hill R, Gardner SE. "There's Not Enough Knowledge Out There": Examining Older Adults' Perceptions of Digital Technology Use and Digital Inclusion Classes. J Appl Gerontol 2019 Aug;38(8):1147-1166. [doi: 10.1177/0733464817737621] [Medline: 29165038]
48. Shapira N, Barak A, Gal I. Promoting older adults' well-being through Internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [doi: 10.1080/13607860601086546] [Medline: 17882585]
49. Seifert A, Cotten SR, Xie B. A Double Burden of Exclusion? Digital and Social Exclusion of Older Adults in Times of COVID-19. Corrected proof. J Gerontol B Psychol Sci Soc Sci 2020 Jul 16:gbaa098 [FREE Full text] [doi: 10.1093/geronb/gbaa098] [Medline: 32672332]

Edited by G Eysenbach; submitted 22.11.20; peer-reviewed by M Baez, A Seifert; comments to author 07.12.20; revised version received 27.12.20; accepted 29.12.20; published 22.01.21.

Please cite as:
Cohen-Mansfield J, Muff A, Meschiany G, Lev-Ari S
Adequacy of Web-Based Activities as a Substitute for In-Person Activities for Older Persons During the COVID-19 Pandemic: Survey Study
J Med Internet Res 2021;23(1):e25848
URL: http://www.jmir.org/2021/1/e25848/
doi:10.2196/25848
PMID:33439851

Original Paper

Expanding Telemonitoring in a Virtual World: A Case Study of the Expansion of a Heart Failure Telemonitoring Program During the COVID-19 Pandemic

Sahr Wali[1,2*], MSc; Milena Guessi Margarido[1,2,3*], PhD; Amika Shah[1,2], MScCH; Patrick Ware[1], MPH, PhD; Michael McDonald[4,5], MD; Mary O'Sullivan[4], RN; Juan Duero Posada[5,6], MD, MSc; Heather Ross[4,5,7], MD, MHSc, FRCPC; Emily Seto[1,2], PEng, PhD

[1]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Department of Computer Systems, Institute of Mathematics and Computer Science, University of São Paulo, São Paulo, Brazil

[4]Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada

[5]Ted Rogers Centre for Heart Research, University Health Network, Toronto, ON, Canada

[6]Advanced Heart Failure, Transplantation and Mechanical Circulatory Support, Toronto Western Hospital, University Health Network, Toronto, ON, Canada

[7]Department of Medicine, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Sahr Wali, MSc
Centre for Global eHealth Innovation
Techna Institute
University Health Network
TGH/RFE Building, 4th Floor
190 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 4800 ext 4765
Email: sahr.wali@mail.utoronto.ca

Abstract

Background: To minimize the spread and risk of a COVID-19 outbreak, societal norms have been challenged with respect to how essential services are delivered. With pressures to reduce the number of in-person ambulatory visits, innovative models of telemonitoring have been used during the pandemic as a necessary alternative to support access to care for patients with chronic conditions. The pandemic has led health care organizations to consider the adoption of telemonitoring interventions for the first time, while others have seen existing programs rapidly expand.

Objective: At the Toronto General Hospital in Ontario, Canada, the rapid expansion of a telemonitoring program began on March 9, 2020, in response to COVID-19. The objective of this study was to understand the experiences related to the expanded role of a telemonitoring program under the changing conditions of the pandemic.

Methods: A single-case qualitative study was conducted with 3 embedded units of analysis. Semistructured interviews probed the experiences of patients, clinicians, and program staff from the Medly telemonitoring program at a heart function clinic in Toronto, Canada. Data were analyzed using inductive thematic analysis as well as Eakin and Gladstone's value-adding approach to enhance the analytic interpretation of the study findings.

Results: A total of 29 participants were interviewed, including patients (n=16), clinicians (n=9), and operational staff (n=4). Four themes were identified: (1) providing care continuity through telemonitoring; (2) adapting telemonitoring operations for a more virtual health care system; (3) confronting virtual workflow challenges; and (4) fostering a meaningful patient-provider relationship. Beyond supporting virtual visits, the program's ability to provide a more comprehensive picture of the patient's health was valued. However, issues relating to the lack of system integration and alert-driven interactions jeopardized the perceived sustainability of the program.

Conclusions: With the reduction of in-person visits during the pandemic, virtual services such as telemonitoring have demonstrated significant value. Based on our study findings, we offer recommendations to proactively adapt and scale telemonitoring programs under the changing conditions of an increasingly virtual health care system. These include revisiting the scope and expectations of telemedicine interventions, streamlining virtual patient onboarding processes, and personalizing the collection of patient information to build a stronger virtual relationship and a more holistic assessment of patient well-being.

(J Med Internet Res 2021;23(1):e26165) doi:10.2196/26165

KEYWORDS

telemedicine; telehealth; digital health; digital medicine; virtual care; COVID-19; coronavirus; SARS-CoV-2; public health; surveillance; outbreak; pandemic; infectious disease; cardiology; patient

Introduction

Shifting Societal Views on Essential Health Services

The unprecedented magnitude of the COVID-19 pandemic has shifted the focus of the health care system toward patients infected with SARS-CoV-2 [1]. Through these uncertain times, to minimize the spread and risk of a COVID-19 outbreak, societal norms have been challenged in terms of what health services are considered *essential* [2]. Specifically, health care systems have been faced with a significant challenge whereby in-person visits, surgeries, and treatments once considered *essential* have either been cancelled or postponed to reduce the likelihood of virus transmission [2,3]. With the implementation of these public health measures, patients with chronic conditions relying on traditional models of clinic-based care are now at higher risk of health deterioration [4,5]. These patients have been reported to face a higher risk of severe illness when infected with COVID-19. Thus, maintaining patient health status in the context of the current health care reconfigurations represents a significant challenge [5,6].

Virtualizing Care During the COVID-19 Pandemic

To ensure access to care in light of restrictions to in-person visits, innovative virtual care models have been used during the pandemic as a necessary alternative [4,6]. The value of virtual care has been demonstrated in a number of studies through its ability to provide timely care that enables patients to maintain physical distancing while also conserving health care resources [6,7]. Among virtual care technologies, telemonitoring is being increasingly promoted as a tool to assist patients who need more frequent care touchpoints [7]. In the field of telehealth, telemonitoring has been defined as the use of technology to remotely monitor and transmit data relating to patient health status in geographically separated settings. With features enabling providers to closely monitor patient symptoms and vital parameters at home, the use of telemonitoring has been viewed as a valuable addition to the management of patients [8,9].

Since the pandemic started, large organizations such as the Canadian Cardiovascular Society and the Heart Failure Society of America have recommended transitions to virtual modes of care to meet the needs of patients during the pandemic [4,5]. In New York City, where high transmission of COVID-19 has been reported, telemonitoring was found to be an important component of the medical response to the pandemic because it reduced the demand on strained health care staff and enabled the meeting of patient needs at home [10]. As the health care system began to recognize the new reality and the presence of physical distancing restrictions, potential arose for virtual services such as telemonitoring to play an important role in patient management [3,10]. However, it is important to recognize that many telemonitoring programs were originally designed under the conditions of a pre–COVID-19 model of care that may no longer be available due to the limitations on in-person visits. With this growing shift to increasingly virtual health systems, there is a need to evaluate the functionality of a program under changing circumstances to optimize its use for conditions both during and beyond the pandemic.

The objectives of this paper were to (1) understand how telemonitoring is being used during the pandemic through the rapid expansion of a heart failure telemonitoring program, (2) report the barriers and facilitators related to the new virtual delivery of the telemonitoring program, and (3) evaluate the components of the telemonitoring program that require adaptation to sustain its use during and after the pandemic. In reflection of these objectives, the research question was: *With the changing conditions associated with the COVID-19 pandemic, what are the experiences and perceptions of patients, providers, and staff regarding the use of a telemonitoring program to support care needs?*

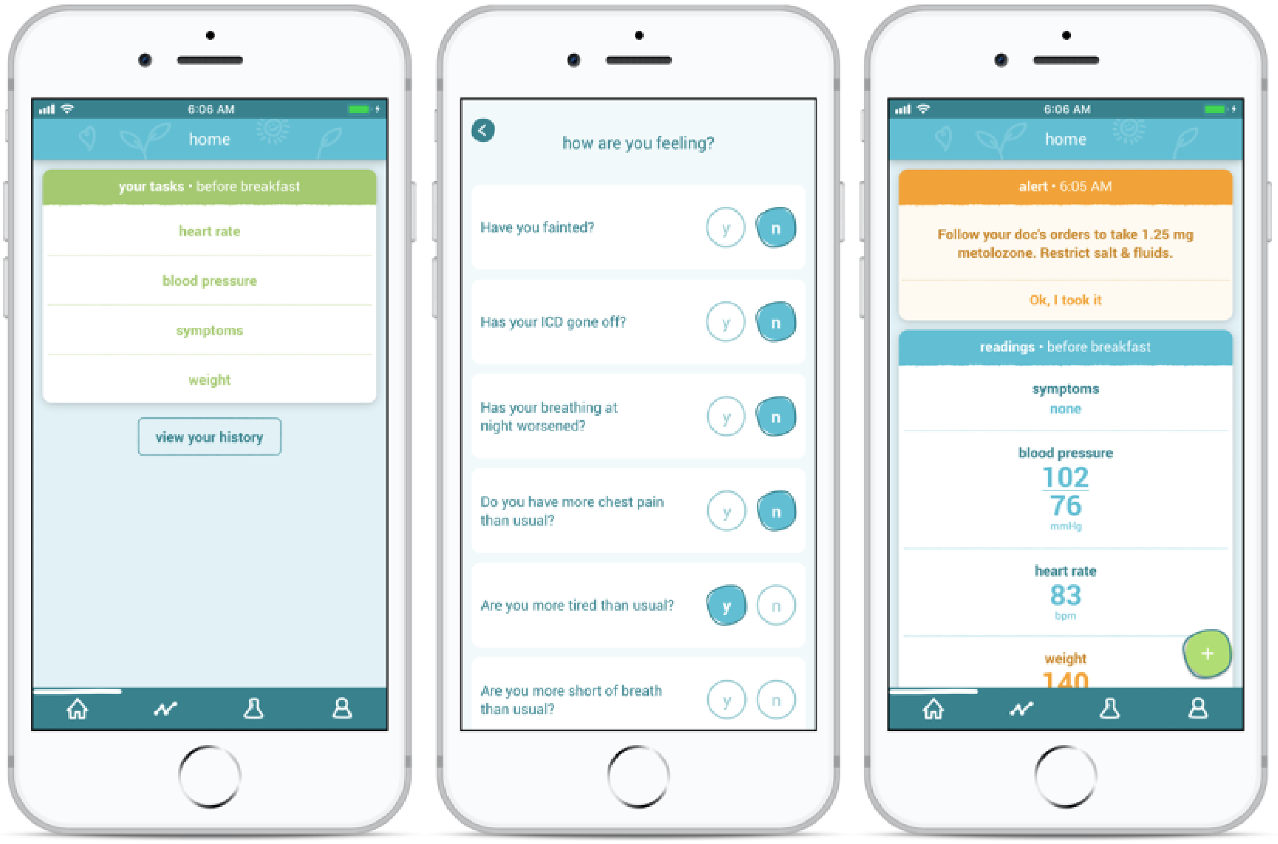
Methods

Setting

With the growing concern regarding COVID-19, the Peter Munk Cardiac Centre (PMCC) Heart Function Clinic at the Toronto General Hospital began to transition many of its services to a virtual care model. The majority of in-person appointments were rescheduled or replaced by videoconferencing visits or telephone calls. Clinicians affiliated with the PMCC Heart Function Clinic also had the option to enroll patients in the existing "Medly" program, a mobile phone–based telemonitoring program designed to provide remote clinical support for patients with heart failure [11,12]. The Medly program currently does not have any strict clinical criteria dictating patient eligibility for enrolment. Instead, cardiologists use their clinical judgement to refer patients they deem would benefit from the program. With Medly, patients use a smartphone with the Medly app, a weight scale, and a blood pressure monitor to record daily physiological readings and symptoms (Figure 1). Using these devices, patients are instructed to take daily weight and blood pressure readings as well as to record their symptoms through a series of self-report questions on the Medly app [12]. A rules-based algorithm uses these data to automatically generate individualized self-care instructions for the patients while simultaneously alerting clinicians when there are changes in the patient’s health status beyond their set clinical parameters [12]. Clinical alerts, delivered via email or the Web-based Medly dashboard, are mostly managed by the Medly Coordinator, a registered nurse within the Heart Function Clinic, with alerts being escalated to the cardiologists as needed. As the Medly system is a class II Health Canada–approved medical device, all data collected from the system are stored in the secure servers at the University Health Network, whereby patient information can only be accessed by the Medly Coordinator and the rest of the patient’s clinical team.

The Medly Coordinator also plays a fundamental role in the onboarding process, where during an in-person clinic visit, the treating cardiologist first presents the program to the patient and gains their approval for enrolment. The Medly Coordinator then meets with the patient to review the program processes (ie, account creation, training, study consent) and to assess their equipment needs [12]. Patients who require all equipment to be provided receive a full Medly kit, whereas other patients follow the bring your own device model and use some or all of their own devices (ie, smartphone, weight scale, blood pressure cuff) [12]. To support the rapid expansion of the Medly program during the COVID-19 pandemic, two nurses from the cardiology department within the Toronto General Hospital were seconded to assist with the foreseen increase in patient enrollment. As of March 8, 2020, the program had previously served 565 patients; however, in preparation for the uncertain impact of the pandemic and the need for remote support, plans were made to enable the Medly program to accommodate a rapid increase of up to 1000 patients. From March 9 to June 8, 2020, a total of 117 new patients were onboarded.

Figure 1. Screenshots of the Medly app displaying the patient self-care tasks, the symptom questionnaire, and the resultant self-care feedback after all tasks are completed and processed by the algorithm.



how are you feeling?

Have you fainted? y n

Has your ICD gone off? y n

Has your breathing at night worsened? y n

Do you have more chest pain than usual? y n

Are you more tired than usual? y n

Are you more short of breath than usual? y n

Follow your doc's orders to take 1.25 mg metolozone. Restrict salt & fluids.

Study Design

This study used a qualitative case study design [13] with the case defined as the telemonitoring program at the Heart Function Clinic at the Toronto General Hospital. The uses of an existing telemonitoring program by patients, clinicians, and operational staff were selected as the units of analysis to better understand the ways in which the pandemic has impacted their use and satisfaction of the Medly program.

Recruitment of Participants

Patients were purposefully identified and recruited according to a range of demographic characteristics, including age, gender, location of residence (rural, suburban, urban), ethnicity, income, and comfort with technology. Although enrolment in the Medly program is at the discretion of each patient’s cardiologist, patients were considered eligible for this study if they spoke English and were users of the Medly program. We aimed to recruit patients who were onboarded to the Medly program both before and during the pandemic. To ensure that the patients onboarded during the pandemic had some experience using the system, only patients enrolled in the program for at least 20 days were considered to be eligible for the study. All 12 clinical staff members and 4 operational staff members of the telemonitoring program were invited to participate. All participants were recruited in alignment with an existing quality improvement study evaluating the telemonitoring program (University Health Network Research Ethics Board 16-5789 and University of Toronto Research Ethics Board 39449). Eligible patient participants were recruited until data saturation was reached, as in, no new perspectives or ideas were generated from the data. Specifically, as the data collection and data analysis processes were conducted on a continual basis, data saturation was deemed complete when no new insights were found from the latest 3 patients interviewed. This included, but was not limited to, patient feedback related to their experience using the program as well as challenges associated with the pandemic.

Data Collection and Analysis

To accommodate physical distancing measures, in-depth semistructured interviews were conducted over the telephone by two authors experienced in qualitative research (AS and SW). Interview guides were developed based on the core components from the Benefits Evaluation Framework using semistructured, open-ended questions [14]. Each interview guide was developed and tailored for the patient, clinician (nurses and cardiologists), and operational staff participant groups. During each interview, participants were asked to comment on their experiences with managing heart failure both before and during the COVID-19 pandemic, as well as their experiences with the Medly telemonitoring program. Interviews were conducted between May 4 and June 18, 2020, and lasted approximately 30 minutes. All interviews were digitally recorded and professionally transcribed verbatim for analysis.

Interview transcripts underwent inductive thematic analysis at the semantic level according to the iterative 6-phase approach outlined by Braun and Clarke in 2006 [15]. Eakin and Gladstone’s “value-adding” approach to qualitative analysis was also used to enrich the analytic interpretation of the study findings and enhance the level of contextualization to the knowledge generated [16]. By using analytic devices such as reflexivity and generative coding to review each transcript, the authors were able to deepen their interpretation of the data in more abstract terms. Three authors were involved in the data analysis process (AS, MG, SW). Throughout the research process, all authors engaged in both procedural and analytical memoing to improve the overall trustworthiness of the analysis [17]. Transcripts and analytic memos were then entered into NVivo 9 (QSR International) to initiate the organization and analysis of the data. To gain a holistic perspective on all the data collected, one author (MG) independently analyzed all interview transcripts, whereas in parallel, two authors independently analyzed either the patient (AS) or clinician and staff (SW) transcripts. All authors subsequently met to discuss and compare codes for each participant group. Following the initial discussions, the codes were then grouped into categories to identify themes relating to the research question. After five analytic discussions, the research team collectively developed four themes. The four core themes were then individually reviewed by the study team to ensure internal coherence, consistency, and approval [15].

Results

Participant Characteristics

A total of 29 participants were interviewed: 16 patients, 9 clinicians (ie, cardiologist, registered nurse), and 4 operational staff. The characteristics of the interviewed patients are shown in Table 1. These data were collected via a self-report questionnaire as part of an existing quality improvement evaluation of the Medly program.

Table 1. Characteristics of the patient interview participants (N=16).

Characteristic	Value
Age (years), mean (SD; range)	54.5 (19.9; 23-78)
Sex, n (%)	
Male	8 (50)
Female	8 (50)
Ethnicity, n (%)	
White	10 (63)
Black	1 (6)
Filipino	1 (6)
South Asian	1 (6)
Southeast Asian	2 (13)
Not declared	1 (6)
Place of birth, n (%)	
Canada	12 (75)
Other	3 (19)
Not declared	1 (6)
Higher education achieved, n (%)	
High school	2 (13)
Trade or technical training	4 (25)
College or university	8 (50)
Postgraduate	1 (6)
Not declared	1 (6)
Rurality, n (%)	
Urban	4 (25)
Suburban	8 (50)
Rural	3 (19)
Not declared	1 (6)
Living arrangement, n (%)	
Living with family/partner	13 (81)
Living alone	2 (13)
Not declared	1 (6)
Income (CAD $),[a] n (%)	
<15,000	4 (25)
15,000-49,999	3 (19)
50,000-74,999	6 (38)
>75,000	1 (6)
Not declared	2 (13)
Time of enrolment , n (%)	
Before March 2020	4 (25)
After March 2020	12 (75)
Participation model , n (%)	
Full kit	2 (13)
Bring your own phone	4 (25)

Characteristic	Value
Bring your own everything	9 (56)
Not declared	1 (6)
Uses a smartphone , n (%)	
Yes	11 (69)
No	3 (19)
Not declared	2 (13)
Frequency of smartphone use , n (%)	
Frequently	8 (50)
Sometimes	5 (31)
Not declared	3 (19)
Comfort with technology , n (%)	
Very comfortable	3 (19)
Somewhat comfortable	2 (13)
Comfortable	4 (25)
Not comfortable	2 (13)
Not declared	5 (31)
New York Heart Association classification , n (%)	
Class I	2 (13)
Class II	9 (56)
Class III	2 (13)
Class IV	0 (0)
Not declared	3 (19)

[a] CAD $1=US $0.79.

Interview Findings

During the analysis of all participant data, four themes were identified: (1) providing care continuity through telemonitoring; (2) adapting telemonitoring operations for a more virtual health care system; (3) confronting virtual workflow challenges; and (4) fostering a meaningful patient-provider relationship.

Providing Care Continuity Through Telemonitoring

With the significant reduction of in-person clinic visits, the use of telemonitoring had become more desirable, as it provided clinically relevant patient data to help support virtual visits or the postponement of a visit. Although there are no strict clinical criteria guiding patient eligibility for Medly beyond requiring the patients to be followed by the Heart Function Clinic, clinicians increased the number of patients they referred to the program during the pandemic. With the limitations created by COVID-19, cardiologists were less selective in patient referrals to Medly, as their previous concerns for onboarding no longer weighed as heavily on their decision-making. Consequently, despite provider concerns, program nurses reported that patients with less familiarity with technology were able to use the application with minimal assistance:

> *One of the determining factors for me to onboard someone on Medly before COVID was if I had concerns about patients not being able to adhere to the daily weights reliably for whatever reason, I would hesitate to on-board them. Now that plays less of a role in my decision-making.* [Cardiologist 3]
>
> *Some of the patients are not always the most tech-savvy but they're trying their best. And a few explanations on the phone and, you know, they're good with it, you know?* [Nurse 1]

For many patients, with the heightened fear of contracting COVID-19 and the lack of accessibility to care services, Medly provided patients with a sense of security and comfort for their care management. New patients onboarded onto Medly appreciated the continual monitoring provided by the telemonitoring program, as it created a clinical safety net during a time when society was left in a state of uncertainty. The convenience and ease of being able to directly connect with a health care provider who was familiar with their medical history was also highlighted as a key value of the program:

> *It would be a wonderful thing for a lot of people in here to have somebody that had to check on them in the morning. I feel it's almost doing [it] to me because my vital signs don't change that much but at least it checks that I'm still alive and breathing.* [Patient 2]
>
> *My patients really liked the opportunity to be enrolled and be involved, they didn't know that it [Medly] existed before because they were sort of lower-risk*

> *patients who I hadn't thought to enroll in Medly before and now that they have some way to send in their vitals and have someone keep track of them a little more closely for a lot of patients, that gives them a lot of added security.* [Cardiologist 1]

Although clinicians were able to connect with patients through other modes of virtual engagement (ie, telephone calls, videoconferencing), they felt that Medly's continual monitoring provided a more comprehensive picture of patient health compared to virtual visits alone. Specifically, clinicians using the Medly system were able to monitor trends in patient heart rate, blood pressure, weight, and symptoms, all of which helped detect worsening health and the potential need for patient follow-up. With the conditions of the pandemic and the absence of in-person visits, the metrics provided by Medly were highlighted to help provide a level of care continuity and context to support clinician decision-making:

> *I think it's driving with your eyes open rather than closed and, to some degree, a lot of what you're trying to do in terms of adjusting medications is to try and keep their weight within range and blood pressure and heart rate and their symptoms under control, and you can get all [of] that from Medly and track their trends over time.* [Cardiologist 2]

Beyond the usual benefits of telemonitoring, clinicians valued that Medly enabled them to make clinical assessments or changes in patient care that would usually be completed in-person. In the context of COVID-19 limitations, treatment optimization was still feasible, whereby clinicians were able to effectively titrate mediation without in-person patient visits to the hospital. Both patients and clinicians felt reassured that the quality of care had not been jeopardized under the shifting conditions of a virtual care model:

> *We've been able to titrate medication remotely, which is probably one of the biggest things because previously we would see patients in the clinic and then we wouldn't change a new med until maybe the next time they were seen or a couple months down the road. I've been keeping track of many patients who are in the titration phase in their medication and we've been able to optimize their medications on a pretty rapid basis, like biweekly, to achieve triple therapy medication optimization, all while doing everything remotely.* [Nurse 3]

Adapting Telemonitoring Operations for a More Virtual Health Care System

With the rapid push for virtual care, many of the Medly program's in-person processes were quickly virtualized to meet COVID-19 safety requirements. Despite common belief that the telemonitoring program was already functioning under a fully remote/virtual model, operational staff highlighted that while the onboarding process previously required patients to be introduced to the program following a clinic visit, these in-person procedures were now converted into a series of telephone calls with the Medly Coordinator ahead of a virtual visit. Clinicians and operational staff reported mixed patient reactions to the remote onboarding process; some reported that patients expressed confusion regarding why they were being enrolled in the program. Overall, the clinicians felt that patients had a more negative reaction in the early weeks of the pandemic, mainly because they had been preemptively contacted by the nurse coordinator for onboarding without receiving a contextual introduction from their cardiologist:

> *I think when patients have been seen in the clinic, in their virtual clinic, and then are enrolled onto Medly, their response is much better because they've already heard about it from their physician. Like, if they had had a clinic in the morning and then if you told them the nurse will call you today and get you set up, I want to remotely monitor you. That response has been really positive.* [Nurse 3]

One member of the operational staff reflected on the benefits of remote onboarding and found that the use of a low-contact setup enabled many patients to participate in the program without needing to leave their home. Specifically, with this new virtualized onboarding process, after the Medly Coordinator identified the type of Medly kit needed (eg, full kit, weight scale only), this equipment was delivered directly to the patient. However, clinicians and operational staff reported that this new process also resulted in delayed patient participation in the program, as some patients were unable to report their measurements due to their equipment still being in transit:

> *And then we've had to, like, mail out a lot of equipment to people who need it, and especially right now because people were not necessarily willing to go out and get their own equipment during COVID.* [Staff 2]
>
> *So, there's a bit of like waiting involved whereas like normally you kind of expect the patients to be taking measures like the next day. In this case, it's like potentially like one to two weeks later and then even then it's kind of like, oh do you still not have your equipment yet or like how is that going?* [Staff 1]

As the pandemic progressed, clinicians and program staff reflected on the current operations of the Medly program and found that a number of patient tasks were more challenging within the context of COVID-19. For instance, many patients were required to obtain blood work for remote titrations; however, they felt that the risk associated with leaving their home outweighed the benefit of their treatment optimization. In addition, when patient symptoms worsened beyond their clinical parameters, clinicians or the automated Medly self-care feedback would direct patients to go to the hospital. However, this self-care advice was found to no longer be suitable, given the fear and heavy concern regarding virus exposure:

> *There are a number [of patients] that were quite concerned about going to get blood work, so they didn't want to leave the house…It is a delicate balance on how much can we do through the phone, through Medly, make changes, but in the end we still need to get some of that data in order to make those clinical decisions.* [Cardiologist 1]
>
> *Patients have been also scared of going to hospitals. Even when we want them to come to the hospital to*

> *be assessed in person or to go to the emergency room many of them have pushed back and said that they'd rather deal with their symptoms at home rather than coming to the emergency room and expose themselves to a potential exposure to COVID-19.* [Nurse 1]

With physical distancing requirements in place, many program staff also noticed that patients residing at home were more often alone, without the presence of a caregiver. This loss of social support led many patients to experience a greater need for technical and emotional assistance that was normally provided by a visiting family member. With this, operational staff indicated that during the introductory period of the program, patients onboarded during the pandemic made more telephone calls requesting technical support than patients onboarded before the pandemic:

> *Some of the elderly patients have difficulty with technology where they haven't been exposed to smartphones and that kind of thing. One of them actually said to me that he would usually get his kid to like show him how to do stuff or set stuff up on his phone, but because of COVID they haven't been able to visit so he hasn't been able to sort of get them to help.* [Staff 1]

Confronting Virtual Workflow Challenges

Clinicians experienced many challenges when using Medly in combination with other dedicated systems (eg, the electronic medical record, laboratory systems). Due to the lack of integration between the various systems, clinicians were required to navigate through multiple technologies to obtain up-to-date patient information. Clinicians felt that reviewing the patient information in a single system would not provide them with sufficient information to make a comprehensive clinic assessment. Thus, increased clinical workload was required to access multiple systems to fill the information gaps in their patient profiles:

> *When [radiology and cardiography] don't appear it's much harder to piece things together. So, when you're seeing a patient, which is where it usually comes up, you want the data in one place. Like the last thing in the world I want to do is, I see a patient and I'm like "OK, I'm going to adjust the meds" –. But you [have] to go out of the EMR [electronic medical record]…then you have to log into the Medly platform which requires another password. You're all doing this with the patient on the phone; it's completely clunky.* [Cardiologist 4]

To accommodate the influx of patients during the rapid expansion, two nurses within the Toronto General hospital cardiology department were seconded to support the Medly program. The additional nurses enabled better balancing of the patient caseload. However, several issues regarding patient profile accessibility were reported. Specifically, the nurses indicated that it was difficult to access information about patients who were not part of their caseload. The nurses strongly desired a way to access information for patients not assigned to them to make it easier when they had to cover for one another:

> *[One nurse] watched through all the alerts and so, with the ramping up of numbers more coordinators coming on, we had to find a way to change the caseloads and find who was watching [which patients]. So, there was a way where they changed it so we could easily identify which coordinator was the one watching which patient.* [Nurse 3]

> *[…] finding ways to look at different patients if they're not your own. They're separated [now], but there [should be] a way to follow up with other patients if you were covering for another nurse and need to look up other patients.* [Nurse 4]

Even with the increase in nursing staff, clinicians perceived an increase in their workload during the pandemic. Cardiologists were responsible for monitoring patient alerts over the weekends; however, responding to alerts became even more challenging with the increase in patient caseload. Ultimately, the clinicians thought that as more alerts would accumulate over the weekend, there could be a delay in the response time to contact the patients compared to during weekdays:

> *I have somewhere between 180 and 200 patients on Medly. So, on the weekend, without the nurse support, there's a little bit more work for me to do.* [Cardiologist 5]

Fostering a Meaningful Patient-Provider Relationship

With the transition to a fully virtual model, patients onboarded during the COVID-19 pandemic felt that there was a lack of clarity regarding their purpose of enrollment into the program. Due to the restrictions associated with the pandemic, the Medly Coordinator was unable to provide patients with the initial face-to-face touchpoint that previous patients had received. Thus, many patients that were previously only seen every 6 months to a year were confused in regard to why they were required to input symptoms on a daily basis or form a relationship with the Medly Coordinator, who was previously not part of their care:

> *[Patients tell me], why do you suddenly want to hear from me every single day when prior to this you only saw me once a year… So, there are no negative responses; just a few people skeptical, "who are you and why are you calling me," if they hadn't been pre-warned by their physician.* [Nurse 1]

Clinicians also expressed that it was challenging to build a relationship with patients who were new to the Medly program. Multiple clinicians felt that they were not able to establish a personal connection or objectively assess their patients' condition over the phone, as there was an absence of visual cues or caregiver support to obtain patient information. Clinicians indicated that patient body language often provided a good depiction of their well-being, and in cases where the patient was unable to vocalize their health issues, a caregiver would often assist in relaying relevant health information. Clinicians were concerned that without face-to-face visits or caregivers present, they would be unable to establish a relationship with patients or assess their condition accordingly:

> *What I have noticed is that, for new patients, it's very hard to establish a good relationship... I called them a couple of times and we couldn't establish a kind of personal relationship. And I think that that was a bit of a problem. I think that that can increase their chances…of non-compliance.* [Cardiologist 2]

Regarding the frequency of interactions needed, clinicians had mixed views about how often they should be contacting patients over the telephone. One clinician found that a greater frequency of patient interactions strengthened the trust and basis of their patient relationship. However, other clinicians noted that most of their patient interactions were triggered by system alerts, and any further touchpoints with their patients would be burdensome to their workload:

> *Because the case load has increased, you do over time get a bit of an understanding for your … like any relationship, you need to kind of understand … what they prefer, what they are going through and how they would react.* [Nurse 3]

For patients, the perception of alert-driven relationships meant that they assumed clinicians would often assess patient well-being by the clinical parameters entered into the telemonitoring system. A lack of feedback from the nurses was usually perceived as a validation that the patient was doing well. However, patients who were not experiencing severe heart failure symptoms, and thus were not generating telemonitoring alerts, were dissatisfied because they were unable to establish a strong connection with their clinician. These individuals felt that many factors that contribute to their health were not being taken into consideration (eg, pain management, sleep, living conditions). With the influx of patients due to the pandemic, this issue around alert-driven relationships was further exacerbated, as clinicians faced an increase to their workload that made patient follow-ups beyond system alerts impractical:

> *If I'm the doctor and I'm sitting there, and I've got 100 people to talk to and I got a lot of things to do and all my numbers are perfect, [and I ask] "how are you feeling?" If [my doctor] was to ask me the question like [that] and I'm saying "well I'm not doing very well because I'm feeling like I've got severe pain and spasms in my hip, and my butt. And you're not addressing them.* [Patient 3]

Discussion

Principal Findings

Virtual care interventions have been implemented to provide solutions to the health delivery challenges posed by the COVID-19 pandemic [2,6]. Despite the push for virtual care and the previous support for its use, few studies have explored the perception of these technologies within the contexts of both before and during the pandemic [4]. In this study, we aimed to understand how changes to a heart failure telemonitoring program in response to the COVID-19 pandemic were perceived. From this study, we found that the expansion of the Medly telemonitoring program enabled clinicians to provide a level of cardiac care that would not be possible without telemonitoring. Our findings identified that to sustain the value of a telemonitoring program, it would be necessary to adapt its operational components according to the contextual circumstances (ie, required physical distancing due to the pandemic).

Prior to the COVID-19 pandemic, most health care encounters between patients and providers occurred face-to-face; thus, the incentive for virtual care was limited. The findings from this study indicate that by removing the safety net of in-person care, the restrictions of the pandemic acted as an enabler of virtual care adoption. We observed that many patients who would normally not be considered ideal candidates for the program were readily onboarded despite concerns regarding their technological or health literacy levels. Patients of varying backgrounds used the program beyond expected levels of success; this finding helped challenge the paradigm regarding the characteristics of an "ideal" program candidate [18]. Often, with these biases, many virtual care interventions fail to serve patient groups that are most in need of support [19]. Our findings suggest that re-evaluating the assumptions of who can benefit from a telemonitoring program may help expand the range of patient groups it serves.

While Medly was able to fill a void in patient care, it is important to recognize that the program team was in a unique position to adapt many of its underlying in-person operations. By virtualizing the onboarding process and the delivery of program equipment, the Medly team was able to accommodate the rapid expansion under the province's physical distancing requirements. Other studies have reported that many heart failure clinics attempted to virtualize in-person practices but were unable to maintain these services due to the lack of infrastructure and clinical support available [4]. However, despite the Medly team's success virtualizing the onboarding process, several systems (ie, Medly automated feedback) and clinician instructions were still deemed inappropriate (eg, laboratory work at a local clinic, advice to go to the hospital) in the context of the pandemic. Due to these challenges, similar studies have recommended the development of communication channels or regular consultations with patients to enable them to voice their concerns [20]. Providing patients with these opportunities would help to re-evaluate program components to meet their changing needs and would further empower them to manage their care.

Despite significant efforts to prepare for the rapid expansion, the lack of system integration and established collaborative processes led to escalation of the level of virtual workflow inefficiencies within the program [21]. While the additional nursing staff helped support the patient caseload, the telemonitoring program was operating under a fragmented ecosystem of virtual technologies (ie, the EMR, the Medly system, videoconferencing) that increased the complexity of clinical practices. With these technical challenges, recent research has recommended shifting traditional clinical workflows to a more digitally assisted pathway [6]. Specifically, to prevent duplicating tasks or spending excessive time navigating different systems, the adoption of dedicated communication channels and tailored workflows across multiple platforms would help to increase clinician collaboration and streamline data access.

While the rapid expansion was able to provide access to care and consistent remote management for a larger group of patients, the model of care did not change with disease severity. The clinical parameters were individualized to the patients; however, some perceived that the program was built on the principle of *problem-centered care* [22]. Many patients did not understand the purpose of their enrollment in the program, especially those who were not producing alerts, and they felt that this lack of clarity jeopardized their virtual relationship. Recent studies have attributed similar findings to the widening of the scope of telemonitoring programs to accommodate the limitations posed by the pandemic without considering the impact of the changes on clinical effectiveness or patient satisfaction [6]. We argue that to sustain telemonitoring adoption, clinicians should begin to clarify the specific objectives of the telemonitoring program to prevent possible misconceptions among patients. In addition, a more holistic assessment of each patient's condition should be explored to enhance patients' perceived value of the intervention. For instance, the investigation of psychosocial outcomes has been encouraged in the context of chronic illnesses [23], as monitoring a broader set of health indicators (eg, pain, sleep, mental health) has been reported to contribute to an objective assessment of a patient's well-being [24].

Recommendations

Patient self-management and remote monitoring tools represent an important avenue to deliver high-quality care in the community [25]. With no clear end in sight for the pandemic, health experts are now warning that physical distancing restrictions and other safety measures may be required for many more months or even years [6,26]. In light of our study findings, we offer four recommendations that can be used to help proactively adapt and scale up telemonitoring programs in the context of the increasingly virtualized health systems.

Revisit Scope and Eligibility for the TM Program

Consult with clinicians, staff, and patients to evaluate the scope and expectations of the program under the conditions of the pandemic and beyond. Reflect on how this scope will impact or influence potential expansion efforts to scale the program to other health care settings (ie, disease of focus, intervention delivery). In line with this new scope, consider revisiting the patients served by the program in terms of age, ethnicity, gender, and geographical location to improve accessibility to more underserved groups [27].

Expand Information Modalities Available to Streamline the Remote Onboarding Process

Create an unsupervised (ie, self-paced) dynamic or static tool demonstration, such as an interactive in-app tutorial, to walk patients through key application features. These modalities should emphasize the purposes and patient-specific benefits of the program to foster a greater sense of connection with their care team (eg, recorded audio or video messages from clinicians, surveys to understand equipment needs and difficulties). Ensure that common troubleshooting scenarios are included in the demonstration (eg, expected response to an alert, how to manage a loss of connectivity) to help standardize the onboarding process for all patients.

Support Efficient Modes of Clinician Communication and Patient Management

Consult with clinicians to develop direct communication channels and pathways to streamline navigation across multiple virtual platforms (ie, synchronized and automatically updated EMR and telemonitoring dashboard data). To facilitate task-sharing, nurses should be given access to all patients' profiles without shifting responsibilities for their assigned patient alerts.

Personalize Frequency and Patient Information Collected

Based on the evolving scope of the program, consider monitoring other health indicators (eg, pain, sleep, depression) to provide a more nuanced assessment of patient well-being in the absence of worsening heart failure symptoms. Consult with patients and clinicians to assess the frequency of touchpoints needed based on individual characteristics that may not be driven by alerts. Patient preferences regarding modes of communication, including video, text message, or audio, should also be incorporated to personalize care delivery.

Limitations

There are a number of limitations to this study. First, among the group of clinicians who agreed to participate, 3 were unable to complete the interview due to scheduling challenges. Second, due to challenges posed by COVID-19, we used telephone-based recruitment and interview processes, which may have biased our patient sample toward individuals who are more comfortable with technology. Third, despite our intent to purposefully recruit patients across a range of demographic characteristics, patients in this study were mainly young, White, suburban, and/or college-educated. We recognize that the population of patients at the Heart Function Clinic is largely White and educated; therefore, our sample of patients for this study is reflective of the clinic population but may not be representative of the broader heart failure population in relation to age, ethnicity, rurality, and education. Finally, the recommendations from this study may not be reflective of all the requirements to adapt and scale telemonitoring programs, as it was based solely on the findings of one telemonitoring program for heart failure self-care. With this, we recognize that the recommendations outlined above may describe a subset of the key requirements to proactively adapt and scale telemonitoring programs under the context of increasingly virtualized health care systems due to the pandemic.

Conclusions

The emergence of the COVID-19 pandemic has required society to re-evaluate virtual services as essential. The removal of traditional safety nets for in-person care has provided an opportunity for virtual services such as telemonitoring to demonstrate their value. In this study, we found that the rapid expansion of a telemonitoring program enabled patients to access cardiac care while maintaining physical distancing measures. Telemonitoring was able to fill a void for consistent patient monitoring, although the metrics provided by the system did not always represent an accurate picture of the patients' well-being. Under the conditions of the pandemic, issues surrounding unintegrated, siloed systems (eg, EMR, laboratory systems) were escalated, as they contributed more to the clinician workload burden. To optimize the use of telemonitoring under the conditions of a "virtual world," we present a series of recommendations to help sustain telemonitoring adoption through improvements in workflow efficiency and personalized care. These recommendations may not be inclusive of all the requirements to support the scalability of telemonitoring interventions; however, they will help serve as a foundational guide to better adapt program design. Ultimately, the intent of these recommendations is to improve care delivery for patients of varying needs and capabilities and to better support more patient-centered care both during and after the pandemic.

Conflicts of Interest

None declared.

References

1. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
2. Madjid M, Safavi-Naeini P, Solomon SD, Vardeny O. Potential Effects of Coronaviruses on the Cardiovascular System: A Review. JAMA Cardiol 2020 Jul 01;5(7):831-840. [doi: 10.1001/jamacardio.2020.1286] [Medline: 32219363]
3. Barghash MH, Pinney SP. Heart Failure in the COVID-19 Pandemic: Where Has All New York's Congestion Gone? J Card Fail 2020 Jun;26(6):477-478 [FREE Full text] [doi: 10.1016/j.cardfail.2020.04.016] [Medline: 32389608]
4. Virani SA, Clarke B, Ducharme A, Ezekowitz JA, Heckman GA, McDonald M, Canadian Heart Failure Society. Optimizing Access to Heart Failure Care in Canada During the COVID-19 Pandemic. Can J Cardiol 2020 Jul;36(7):1148-1151 [FREE Full text] [doi: 10.1016/j.cjca.2020.05.009] [Medline: 32405146]
5. Vandekerckhove P, Vandekerckhove Y, Tavernier R, De Jaegher K, de Mul M. Leveraging User Experience to Improve Video Consultations in a Cardiology Practice During the COVID-19 Pandemic: Initial Insights. J Med Internet Res 2020 Jun 25;22(6):e19771 [FREE Full text] [doi: 10.2196/19771] [Medline: 32519964]
6. Silven AV, Petrus AHJ, Villalobos-Quesada M, Dirikgil E, Oerlemans CR, Landstra CP, et al. Telemonitoring for Patients With COVID-19: Recommendations for Design and Implementation. J Med Internet Res 2020 Sep 02;22(9):e20953 [FREE Full text] [doi: 10.2196/20953] [Medline: 32833660]
7. Khairat S, Meng C, Xu Y, Edson B, Gianforcaro R. Interpreting COVID-19 and Virtual Care Trends: Cohort Study. JMIR Public Health Surveill 2020 Apr 15;6(2):e18811 [FREE Full text] [doi: 10.2196/18811] [Medline: 32252023]
8. van Deursen AJ. Digital Inequality During a Pandemic: Quantitative Study of Differences in COVID-19-Related Internet Uses and Outcomes Among the General Population. J Med Internet Res 2020 Aug 20;22(8):e20073 [FREE Full text] [doi: 10.2196/20073] [Medline: 32750005]
9. Lonergan PE, Washington III SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
10. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient Satisfaction With Telemedicine During the COVID-19 Pandemic: Retrospective Cohort Study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]
11. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. User-Centered Adaptation of an Existing Heart Failure Telemonitoring Program to Ensure Sustainability and Scalability: Qualitative Study. JMIR Cardio 2018 Dec 06;2(2):e11466 [FREE Full text] [doi: 10.2196/11466] [Medline: 31758774]
12. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a Heart Failure Telemonitoring Program Implemented as the Standard of Care in an Outpatient Heart Function Clinic: Pretest-Posttest Pragmatic Study. J Med Internet Res 2020 Feb 08;22(2):e16538 [FREE Full text] [doi: 10.2196/16538] [Medline: 32027309]
13. Yin RK. Case Study Research and Applications Design and Methods. Thousand Oaks, CA: Sage Publications; 2013.
14. Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. Healthc Q 2007;10(1):112-6, 118. [Medline: 17326376]
15. Braun, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
16. Eakin J, Gladstone B. "Value-adding" Analysis: Doing More With Qualitative Data. Int J Qual Methods 2020 Aug 27;19(19):160940692094933-160940692094916. [doi: 10.1177/1609406920949333]
17. Birks M, Chapman Y, Francis K. Memoing in qualitative research. J Res Nurs 2008 Jan;13(1):68-75. [doi: 10.1177/1744987107081254]
18. Poli A, Kelfve S, Motel-Klingebiel A. A research tool for measuring non-participation of older people in research on digital health. BMC Public Health 2019 Nov 08;19(1):1487 [FREE Full text] [doi: 10.1186/s12889-019-7830-x] [Medline: 31703655]

19. Brewer L, Fortuna K, Jones C, Walker R, Hayes S, Patten CA, et al. Back to the Future: Achieving Health Equity Through Health Informatics and Digital Health. JMIR mHealth uHealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
20. Olayiwola JN, Magaña C, Harmon A, Nair S, Esposito E, Harsh C, et al. Telehealth as a Bright Spot of the COVID-19 Pandemic: Recommendations From the Virtual Frontlines (. JMIR Public Health Surveill 2020 Jun 25;6(2):e19045 [FREE Full text] [doi: 10.2196/19045] [Medline: 32479413]
21. Sasangohar F, Bradshaw M, Carlson M, Flack J, Fowler J, Freeland D, et al. Adapting an Outpatient Psychiatric Clinic to Telehealth During the COVID-19 Pandemic: A Practice Perspective. J Med Internet Res 2020 Oct 01;22(10):e22523 [FREE Full text] [doi: 10.2196/22523] [Medline: 32936768]
22. Cowie MR, Chronaki CE, Vardas P. e-Health innovation: time for engagement with the cardiology community. Eur Heart J 2013 Jul 24;34(25):1864-1868. [doi: 10.1093/eurheartj/ehs153] [Medline: 22733834]
23. McAllister M, Dunn G, Payne K, Davies L, Todd C. Patient empowerment: the need to consider it as a measurable patient-reported outcome for chronic conditions. BMC Health Serv Res 2012 Jun 13;12:157 [FREE Full text] [doi: 10.1186/1472-6963-12-157] [Medline: 22694747]
24. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. JAMA Intern Med 2013 Apr 22;173(8):702-704 [FREE Full text] [doi: 10.1001/2013.jamainternmed.61] [Medline: 23460235]
25. World Health Organization, Organisation for Economic Co-operation and Development, International Bank for Reconstruction and Development. Delivering quality health services: a global imperative for universal health coverage. World Health Organization, OECD & International Bank for Reconstruction and Development. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/272465/9789241513906-eng.pdf?sequence=1&isAllowed=y [accessed 2020-08-30]
26. De Simone V, Guarise P, Guardalben S, Padovani N, Tondelli S, Sandrini D, et al. Telecardiology during the Covid-19 pandemic: past mistakes and future hopes. Am J Cardiovasc Dis 2020;10(2):34-47 [FREE Full text] [Medline: 32685262]
27. Were MC, Sinha C, Catalani C. A systematic approach to equity assessment for digital health interventions: case example of mobile personal health records. J Am Med Inform Assoc 2019 Aug 01;26(8-9):884-890 [FREE Full text] [doi: 10.1093/jamia/ocz071] [Medline: 31188438]

Abbreviations

EMR: electronic medical record
PMCC: Peter Munk Cardiac Centre

Edited by G Eysenbach; submitted 30.11.20; peer-reviewed by P Tripathi, M Ariyan, J Manning, F Tabei; comments to author 21.12.20; revised version received 28.12.20; accepted 13.01.21; published 22.01.21.

Please cite as:
Wali S, Guessi Margarido M, Shah A, Ware P, McDonald M, O'Sullivan M, Duero Posada J, Ross H, Seto E
Expanding Telemonitoring in a Virtual World: A Case Study of the Expansion of a Heart Failure Telemonitoring Program During the COVID-19 Pandemic
J Med Internet Res 2021;23(1):e26165
URL: http://www.jmir.org/2021/1/e26165/
doi:10.2196/26165
PMID:33444153

Original Paper

Impact of COVID-19 on Physical Activity Among 10,000 Steps Members and Engagement With the Program in Australia: Prospective Study

Quyen G To[1], PhD; Mitch J Duncan[2], PhD; Anetta Van Itallie[1], BHMSc; Corneel Vandelanotte[1], PhD

[1]Appleton Institute, School of Health, Medical and Applied Sciences, Central Queensland University, Rockhampton, Australia

[2]Priority Research Centre for Physical Activity and Nutrition, School of Medicine and Public Health, The University of Newcastle, Newcastle, Australia

Corresponding Author:
Quyen G To, PhD
Appleton Institute
School of Health, Medical and Applied Sciences
Central Queensland University
554-700 Yaamba Rd
Rockhampton, 4701
Australia
Phone: 61 7 4930 6456
Email: q.to@cqu.edu.au

Abstract

Background: Physical activity is an important health behavior, due to its association with many physical and mental health conditions. During distressing events, such as the COVID-19 pandemic, there is a concern that physical activity levels may be negatively impacted. However, recent studies have shown inconsistent results. Additionally, there is a lack of studies in Australia on this topic.

Objective: The aim of this study is to investigate changes in physical activity reported through the 10,000 Steps program and changes in engagement with the program during the COVID-19 pandemic.

Methods: Data between January 1, 2018, and June 30, 2020, from registered members of the 10,000 Steps program, which included 3,548,825 days with step data, were used. The number of daily steps were logged manually by the members or synced automatically from their activity trackers connected to the program. Measures on program usage were the number of new registered members per day, the number of newly registered organizations per day, the number of steps logged per day, and the number of step entries per day. Key dates used for comparison were as follows: the first case with symptoms in Wuhan, China; the first case reported in Australia; the implementation of a 14-day ban for noncitizens arriving in Australia from China; the start of the lockdown in Australia; and the relaxing of restrictions by the Australian Government. Wilcoxon signed-rank tests were used to test for significant differences in number of steps between subgroups, between engagement measures in 2019 versus 2020, and before and after an event.

Results: A decrease in steps was observed after the first case in Australia was reported (1.5%; P=.02) and after the start of the lockdown (3.4%; P<.001). At the time that the relaxing of restrictions started, the steps had already recovered from the lockdown. Additionally, the trends were consistent across genders and age groups. New South Wales, Australian Capital Territory, and Victoria had the greatest step reductions, with decreases of 7.0% (P<.001), 6.2% (P=.02), and 4.7% (P<.001), respectively. During the lockdown, the use of the program increased steeply. On the peak day, there were more than 9000 step entries per day, with nearly 100 million steps logged per day; in addition, more than 450 new users and more than 15 new organizations registered per day, although the numbers decreased quickly when restrictions were relaxed. On average per day, there were about 55 new registered users (P<.001), 2 new organizations (P<.001), 25.6 million steps (P<.001), and 2672 log entries (P<.001) more in 2020 compared to the same period in 2019.

Conclusions: The pandemic has had negative effects on steps among Australians across age groups and genders. However, the effect was relatively small, with steps recovering quickly after the lockdown. There was a large increase in program usage during the pandemic, which might help minimize the health impact of the lockdown and confirms the important role of physical activity programs during times of distress and lockdowns.

(J Med Internet Res 2021;23(1):e23946) doi:10.2196/23946

KEYWORDS

exercise; pandemic; lockdown; eHealth; physical activity; COVID-19; engagement; behavior

Introduction

Physical activity is an important health behavior due to its association with many physical and mental health conditions, such as cardiovascular diseases, cancers, diabetes, depression, and anxiety [1-3]. It is recommended that people should engage in at least 150 minutes of moderate to vigorous physical activity per week [1]. During the COVID-19 pandemic, nearly all aspects of society were impacted. Lockdowns with social isolation measures may result in higher levels of stress and a decline in physical activity [4]. However, it may also mean that time saved from less social gatherings and commuting to work could be used to be more physically active. As physical activity improves mental health [1], it may also be used as a coping mechanism against higher levels of stress and anxiety during the pandemic. In addition, people may engage in physical activity more frequently as it was one of a few accepted reasons for leaving home during the March 2019 lockdown in Australia.

Recent studies have shown inconsistent results regarding changes in physical activity during this pandemic. A report by Fitbit with data from 30 million users worldwide showed a decrease in steps across many countries, including a 4% reduction among Australians [5]. However, the analysis only focused on data during the week of March 22, 2020, and effects of other key events on steps were not reported. In addition, detailed analysis was either not conducted or not reported for Australians. Similarly, a worldwide decrease in mean steps of 5.5% (287 steps) within 10 days and 27.3% (1432 steps) within 30 days of the pandemic declaration was also found among 455,404 smartphone users from 187 countries [6]. Again, very limited information was reported for Australia.

In contrast, one survey among 12,913 participants from 139 countries between March 24 and March 30, 2020, found an increase in frequency of exercising by 88% among those who normally exercise 1 to 2 times per week and 8% among those who normally exercise 3 times per week, but a decrease of 14% among those who normally exercise 4 or more times per week [7]. There were also reports that online searches for topics relating to exercise dramatically increased in Australia, the United Kingdom, and the United States starting on March 23, 2020, when lockdowns started in Australia and the United Kingdom [8]. Data among 50,000 subscribers to the WHOOP platform also showed an increase of 1.1% in exercise frequency and 1.8% in time spent on higher-intensity exercise [9]. Finally, a survey among 1491 Australians found that 49% self-reported a decrease in physical activity, whereas 20.7% reported an increase, during the pandemic, indicating that the impact of COVID-19 may not be consistent among population groups with different demographic characteristics [10].

Given inconsistent findings and lack of detailed studies on this topic in Australia, this study used data from the 10,000 Steps Australia program to investigate (1) changes in physical activity reported through the 10,000 Steps platform during the COVID-19 pandemic and (2) changes in the engagement with the 10,000 Steps program during the COVID-19 pandemic. These findings will improve our understanding of the effect of distressing events on physical activity and engagement in physical activity programs and, therefore, inform health policies and tailored intervention design to better deal with future crises.

Methods

Data Sources

The 10,000 Steps program is a web- and mobile-based physical activity promotion program that is funded by the Queensland State Government [11]. To date, the program has registered over 463,000 members (ie, about 3000 new registrations per month), with a total of more than 237 billion steps logged (ie, about 40 million per day). Members register an account with the program website, set a daily step goal, use activity trackers (eg, pedometers, Fitbit, and Garmin) to self-monitor their progress, and participate in challenges to stay motivated. A 10,000 Steps smartphone app is also available for both Apple (iOS) and Google (Android). Any data recorded through the app automatically syncs with the 10,000 Steps website. A detailed description of the program has been published elsewhere [11].

This study used data from new and existing registered members of the program between January 1, 2018, and June 30, 2020, which included 3,548,825 days with step data. Step counts of less than 1000 indicate that the trackers were not worn all day, as even very sedentary people would be expected to accrue more than a 1000 steps a day; in addition, step counts of more than 40,000 are considered as extremely high (eg, technology bug, entry errors, or overreporting) [12,13]. As a result, days with logged steps that were fewer than 1000 steps per day (77,311 days) or more than 40,000 steps per day (10,802 days) were considered invalid and excluded. Data from people not living in Australia (399,711 days), aged less than 18 years or above 80 years (52,551 days), or having less than 7 days of logged data (12,833 days) were also excluded. As a result, 2,995,617 days (84.4%) of data between January 1, 2018, and June 30, 2020, were included in the analysis.

Measures

Demographic characteristics, including date of birth, gender, and state or territory, were self-reported by the members at the time of registration. Age was calculated by subtracting date of birth from June 30, 2020, and converting to years. Age was then dichotomized into two groups based on the average age: 18-45 years and >45-80 years. Location referred to states and territories in Australia, including, New South Wales, Victoria, Queensland, Northern Territory, Western Australia, South Australia, Tasmania, and Australian Capital Territory.

The number of daily steps was either logged manually by the members (45% through the website and 25% through the app) or synced automatically (30%) from the activity trackers connected to the program. As the 10,000 Steps website and app only extract steps from those using activity trackers (ie, Fitbit and Garmin), and given that many members use mechanical pedometers that only track steps, moderate to vigorous physical activity was not used in the analysis. Other measures used to represent the engagement with the 10,000 Steps program were the number of new registered members per day, the number of newly registered organizations per day, the total number of logged steps per day, and the number of step entries per day [14,15].

Key event dates were selected based on their potential influence on physical activity. The list of COVID-19-related events in Australia is presented in Table 1. The nationwide lockdown was imposed about two months after the first case was reported in Australia at the end of January 2020. During the lockdown, in addition to social distancing guidelines [16], nonessential businesses, such as gyms, indoor sports facilities, and clubs, were closed [17]. People were allowed to be outside only for exercise or other essential needs. Restaurants and cafes only offered takeaway and delivery services.

Table 1. Key event dates related to the COVID-19 pandemic.

Date	Description
December 1, 2019	First case with symptoms in Wuhan, China
January 25, 2020	First case in Australia reported
February 5, 2020	A 14-day ban for noncitizens arriving in Australia from China implemented
March 2, 2020	Lockdown starts in Australia
May 8, 2020	Australian Government starts relaxing restrictions

Data Analyses

Python, version 3.7.6 (Python Software Foundation), was used to process and analyze the data. Mean steps were calculated across users for each day. To smooth out daily fluctuations, a 7-day moving average for daily mean steps was calculated and used to create a time series plot showing the trend in steps over time for the entire sample and for each age group and gender. Key event dates were also marked on the plot. Average steps of 7 and 30 days before and after each event date and percentage of change were calculated and used to indicate the impact of an event on steps. The percentage of change in steps 7 and 30 days before and after the event were reported separately for each age group and gender, as well as for each state and territory. Data on the use of the 10,000 Steps program, including the number of new registered users per day, the number of newly registered organizations per day, the total number of logged steps per day, and the number of step entries per day, were also plotted using 7-day moving averages. Wilcoxon signed-rank tests were used to test for significant differences in steps between subgroups, including men versus women, those 18-45 years of age versus >45-80 years of age, and engagement measures in 2019 versus those in 2020 (ie, the number of new registered users, new registered organizations, total logged steps, and step entries). Significant differences in steps before and after an event as well as before and after the lockdown in each state were also tested using Wilcoxon signed-rank tests. All *P* values were two sided and considered significant if <.05.

Results

Of 60,560 members logging step data for at least 7 days, 59.7% (n=36,165) were aged between 18 and 45 years and 67.0% (n=40,583) were women. Of 2,995,617 days with step data that were logged between January 1, 2018, and June 30, 2020, about 63.4% (n=1,898,352) were provided by women and 53.1% (n=1,592,021) were provided by those aged between 18 and 45 years.

Figure 1 shows the average number of steps per day, zooming in on the period between July 1, 2019, and June 30, 2020. In general, there was a downward trend from the peak at the beginning of December 2019, and the bottom was reached in early April 2020. Since then, an increasing trend was observed. However, the difference in 7-day moving averages between days with the highest and lowest steps was small (ie, <1500 steps). A decrease in steps was also observed after the following events: the first case with symptoms in Wuhan, the first case in Australia reported, and the start of the lockdown in Australia. An increase was observed after a 14-day ban for noncitizens arriving in Australia from China. The steps at the time the Australian Government started relaxing restrictions appeared to have already returned back to the level before the lockdown. With regard to the subgroups, men and those above 45 years of age logged more daily steps than women (difference=505 steps; *P*<.001) and younger adults (difference=930 steps; *P*<.001). However, the trends were consistent for the subgroups and indicated that the effects of COVID-19 were not different for genders and age groups (see Figure 2).

Figure 1. Daily mean steps over time and key COVID-19 pandemic events. (1) First case with symptoms in Wuhan, China; (2) First case in Australia reported; (3) A 14-day ban for noncitizens arriving in Australia from China implemented; (4) Lockdown starts in Australia; (5) Australian Government starts relaxing restrictions.

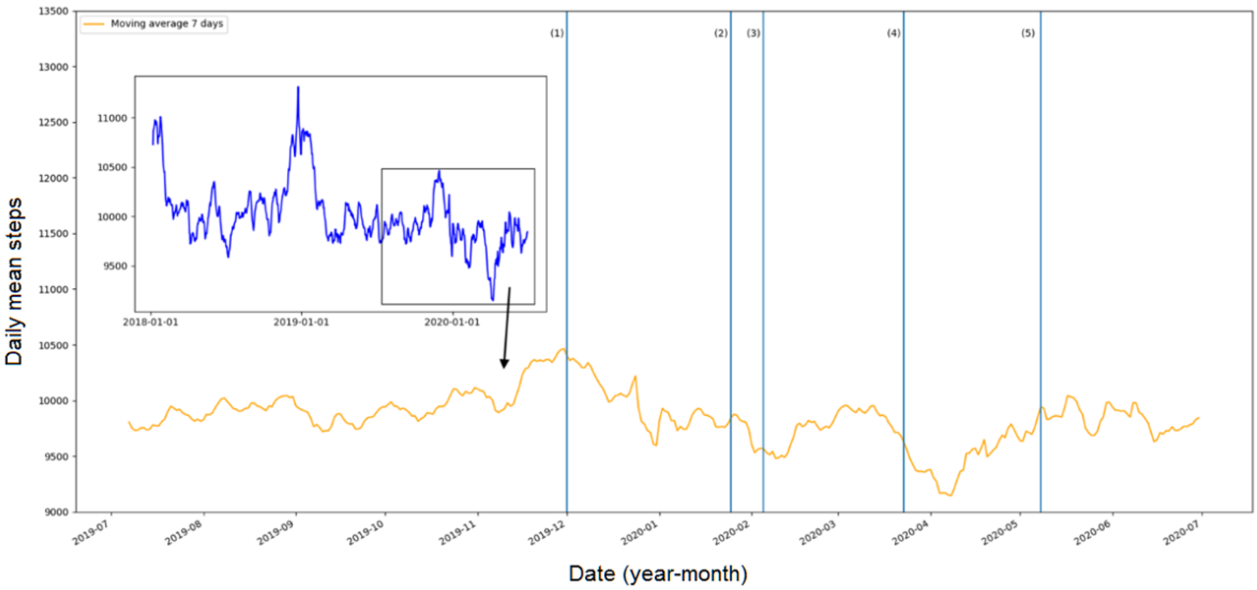


Figure 2. Daily mean steps for subgroups over time and key COVID-19 pandemic events. (1) First case with symptoms in Wuhan, China; (2) First case in Australia reported; (3) A 14-day ban for noncitizens arriving in Australia from China implemented; (4) Lockdown starts in Australia; (5) Australian Government starts relaxing restrictions.

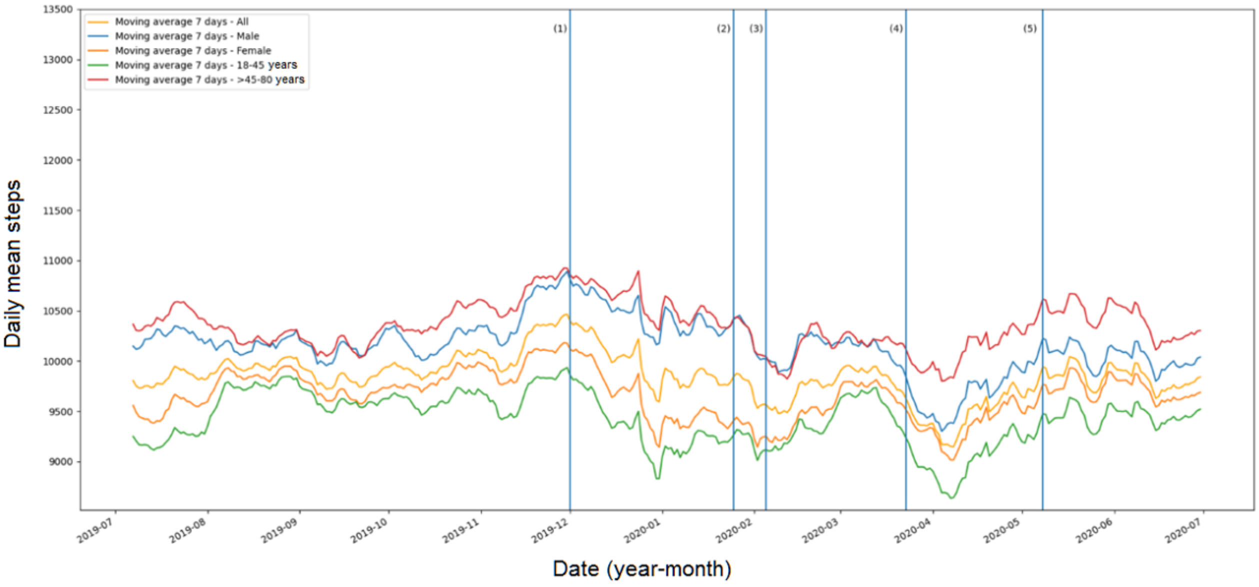


Table 2 shows the average steps of 7 and 30 days before and after an event and for each subgroup. In general, the effects were small for the key events regardless of age group and gender. After the first case was reported in Australia, there was a decrease of 1.5% (*P*=.02) in step average 7 days before and after this date. However, the event with the biggest negative impact on steps was the lockdown starting at the end of March 2020. The reduction was observed for average steps of 7 days (3.4%; *P*<.001) and 30 days (5.0%; *P*<.001) before and after the lockdown. It is also worth noting that the negative impact of a lockdown on steps appeared to be about two times (ie, average of 30 days) larger among those aged between 18 and 45 years compared to older adults. The announcement by the Australian Government to start relaxing restrictions saw a small increase in the step average of 7 days (1.3%; *P*<.001) and 30 days (3.8%; *P*<.001).

Table 2. Average daily steps of 7 and 30 days before and after each event.

Event and participants	Step average of 7 days				Step average of 30 days			
	Before, n	After, n	Difference, %	*P* value	Before, n	After, n	Difference, %	*P* value
First case with symptoms in Wuhan, China								
All	10,292	10,193	–1.0	.26	10,238	10,064	–1.7	<.001
Male	10,707	10,566	–1.3	.22	10,615	10,454	–1.5	.01
Female	10,011	9937	–0.7	.58	9982	9799	–1.8	<.001
18-45 years	9737	9591	–1.5	.06	9676	9479	–2.0	<.001
>45-80 years	10,759	10,701	–0.5	.81	10,718	10,565	–1.4	.004
First case in Australia reported								
All	9778	9634	–1.5	.02	9623	9584	–0.4	.10
Male	10,377	10,201	–1.7	.09	10,204	10,088	–1.1	.02
Female	9331	9208	–1.3	.10	9193	9213	0.2	.89
18-45 years	9219	9117	–1.1	.66	8997	9081	0.9	.25
>45-80 years	10,358	10,169	–1.8	.003	10,289	10,119	–1.7	<.001
A 14-day ban for noncitizens arriving in Australia from China implemented								
All	9462	9393	–0.7	.02	9452	9465	0.1	.90
Male	9873	9745	–1.3	.06	9938	9909	–0.3	.41
Female	9192	9165	–0.3	.16	9134	9175	0.4	.59
18-45 years	9014	9040	0.3	.80	8955	9057	1.1	.11
>45-80 years	9950	9780	–1.7	.002	9999	9913	–0.9	.06
Lockdown starts in Australia								
All	9500	9175	–3.4	<.001	9684	9199	–5.0	<.001
Male	9716	9270	–4.6	<.001	9938	9381	–5.6	<.001
Female	9391	9130	–2.8	<.001	9555	9112	–4.6	<.001
18-45 years	9113	8739	–4.1	<.001	9367	8758	–6.5	<.001
>45-80 years	10,001	9739	–2.6	<.001	10,096	9773	–3.2	<.001
Australian Government starts relaxing restrictions								
All	9637	9767	1.3	<.001	9477	9833	3.8	<.001
Male	9887	10,035	1.5	.004	9723	10,137	4.3	<.001
Female	9489	9606	1.2	<.001	9330	9649	3.4	<.001
18-45 years	9181	9335	1.7	<.001	9060	9426	4.0	<.001
>45-80 years	10,322	10,415	0.9	.02	10,106	10,446	3.4	<.001

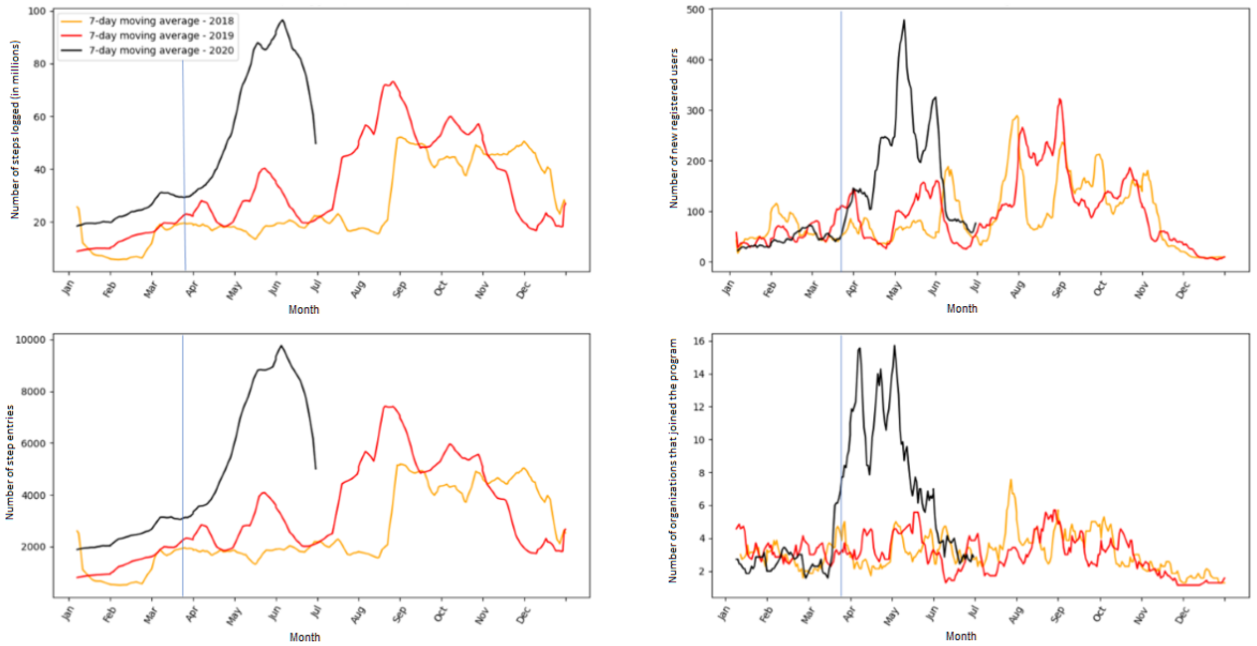
Table 3 shows differences in step averages of 7 days and 30 days before and after the start of the lockdown for each state and territory. A significant decrease was observed in both 7-day and 30-day averages, respectively, for New South Wales (7.0%; *P*<.001 and 5.3%; *P*<.001), Victoria (4.7%; *P*<.001 and 8.1%; *P*<.001), Queensland (1.3%; *P*<.001 and 3.4%; *P*<.001), South Australia (3.2%; *P*=.04 and 2.1%; *P*=.01), Tasmania (3.2%; *P*=.01 and 8.3%; *P*<.001), and Australian Capital Territory (6.2%; *P*=.02 and 10.0%; *P*<.001). The largest decrease for 7 days was seen in New South Wales (about 7%) and the largest decrease for 30 days was seen in Australian Capital Territory (about 10%). No significant difference was found for either Northern Territory (*P*=.15) or Western Australia (*P*=.36).

Table 3. Differences in step averages of 7 and 30 days before and after the start of the lockdown for each state.

State	Step average of 7 days				Step average of 30 days			
	Before, n	After, n	Difference, %	*P* value	Before, n	After, n	Difference, %	*P* value
New South Wales	9343	8686	–7.0	<.001	9528	9021	–5.3	<.001
Victoria	9671	9213	–4.7	<.001	9960	9150	–8.1	<.001
Queensland	9216	9098	–1.3	<.001	9399	9084	–3.4	<.001
Northern Territory	9475	8279	–12.6	.15	9744	8969	–8.0	.07
Western Australia	10,519	10,396	–1.2	.36	10,416	10,424	0.1	.69
South Australia	9914	9596	–3.2	.04	10,028	9816	–2.1	.01
Tasmania	9859	9539	–3.2	.01	10,059	9222	–8.3	<.001
Australian Capital Territory	9545	8955	–6.2	.02	9540	8590	–10.0	<.001

On the other hand, the use of the 10,000 Steps program increased during the pandemic (see Figure 3). The number of steps logged per day and the number of step entries increased steeply since the start of the lockdown. At the same time, the number of new registered users and organizations that joined the program also increased sharply. On the peak day, there were more than 9000 step entries with nearly 100 million steps logged per day; in addition, more than 450 new registered users and 15 new organizations registered per day. However, a sharp decrease in the number of logged steps, step entries, and new registered users and organizations was observed since June 2020. On average per day, there were about 55 new registered users (*P*<.001), 2 new registered organizations (*P*<.001), 25.6 million steps (*P*<.001), and 2672 log entries (*P*<.001) more in 2020 compared to the same period in 2019.

Figure 3. The total number of steps logged, step entries, and new registered users and organizations over time. The vertical line in each graph indicates the start of the lockdown.



Discussion

Principal Findings

This study investigated changes in physical activity among the Australian 10,000 Steps members and the use of the program during the COVID-19 pandemic. The results show that, among the key events, the largest decrease in average steps occurred after the lockdown started in Australia. The finding is consistent with the Fitbit report [5] stating that a reduction of 3% to 4% in steps was observed among the participants in Australia during the lockdown. Although the difference between days with the highest and lowest steps was less than 1500 steps, it has been shown that risk reductions in all-cause mortality are between 6% and 36% for every 1000 steps per day increase [18]. However, it is also worth noting that the decrease in steps was only for several months, and the physical activity level among the participants was already high with average daily steps greater than 9000, which are above the national average for Australia [19]. There also seemed to be a decrease in steps even before the lockdown. This may be because the first case of COVID-19 in Australia was reported at the end of January 2020 and the number of cases increased rapidly well before the lockdown in

March 2020. This was very prominently reported by the media, so it is likely that many people were already adjusting their behavior before they were imposed to do so by the government. Compared to many other countries that have seen much larger decreases in physical activity [6], the effect during the lockdown was small for Australia. This could be attributed to the fact that Australians were still allowed to exercise outdoors during the pandemic, whereas people from many other countries were not. In addition, at the time of data collection, the Australian Government's response to COVID-19, ranked third worldwide, was effective not only in keeping infection and death rates low but also in minimizing social disruption compared to many other countries [20]. As a result, people may have felt less worried and were willing to exercise outside. It is also likely that an increase in steps in April 2020 before the end of the lockdown indicated a partial return to normal life as restrictions eased.

The findings also indicate that technology-based (ie, web and app) physical activity promotion programs, such as the 10,000 Steps program, may play a significant role in minimizing the effect of the COVID-19 pandemic on physical activity. As demonstrated in this study, the number of newly registered users reached a record high during the pandemic, along with billions of steps logged monthly. The engagement with the 10,000 Steps program was much higher during the pandemic compared to earlier periods. Consistent with another study reporting an increase in interest in physical activity at the population level [8], our findings support the notion that people were actively searching for alternative ways to stay motivated to engage in physical activity during the pandemic. A possible explanation may be that people were more often choosing to go for a walk as a way to take a break from work or the restrictions of staying at home. It could also be that physical activity was used as a method to cope with stress and anxiety, as physical activity helps reduce stress and improve mental health [1]. A recent study also found that people who self-reported a decrease in physical activity during the pandemic were likely to experience higher stress and depression [10]. As technology-based physical activity programs are able to provide a social platform to support and encourage people to be more active, even when social distancing is mandated, their role is important, especially during distressing times.

In general, the trend in average steps was consistent across age groups and genders during the pandemic. However, the effect of the lockdown appeared to be larger for younger adults. It is likely due to differences in physical activity preferences between the age groups. As older adults are less likely to participate in team-based activities [21] and are more likely to be active at home or go for a walk around their neighborhood compared to using gyms and sporting facilities [22], the effect may be less severe compared to younger adults. In addition, studies have found that older adults were better at emotion regulation and more likely to use positive reappraisal [23,24] (ie, think positively in a negative situation) as a coping strategy [25], which may help them adjust and maintain their lifestyle behaviors, including physical activity in accordance with social isolation measures. It may also be due to the presence of children in younger adults' households, as studies have shown that parents are likely to be less active compared to nonparents [26,27]. The larger lockdown effect on men could be because men participate more often in competitive physical activities [28], which were more likely to be affected during the pandemic. However, the actual reasons for the observed differences are unclear and more research on this topic is needed.

The immediate effect of the lockdown on step counts was found for most states and territories, with the largest decrease of 7% in New South Wales where a majority of the cases were located in March 2020 [29]. Others with large step count decreases were Victoria, where the second highest number of cases were located, and Australian Capital Territory. Although the number of cases was the third highest in Queensland, the step count decrease was quite small. It is not clear what factors helped, but there was a sharp increase in the use of the 10,000 Steps program, which is based in Queensland. In addition, engagement activities, such as the Queensland Billion Steps Challenge, were implemented by both the government and local organizations. Queensland was also quick in closing the border with New South Wales right after the announcement of the lockdown.

This prospective study included a large number of participants across Australia. Although the main focus was to see the immediate changes in step counts due to COVID-19 events, using several years of data provided additional information on seasonal trends and, therefore, strengthened the results. Time series data also allow examination of both increases and decreases in steps around the key events. However, this study has several limitations. Firstly, many types of physical activity that are dependent on infrastructure that would have closed during the lockdown (eg, swimming pools) were not accounted for by only examining step counts. Secondly, the analysis did not account for differences in lockdown measures between states and territories. Finally, there is a possibility of selection bias, as participation was voluntary. However, the threat of selection bias to validity of this study is minimal due to the fact that our main objective was to see changes in steps over time during the pandemic. When the samples are similar at different time points, the findings will be valid. Furthermore, the sample was large; although a large sample does not mean it is a representative sample, this sample included people from different areas in all states and territories in Australia. Additionally, the 10,000 Steps program supports a variety of step tracking methods and allows both manual entry of steps tracked on any pedometer, activity tracker, or smartphone and automatic syncing of Fitbit and Garmin devices. As such, the risk of selection bias due to the need for an activity tracker or instalment of a specific app was likely low.

Conclusions

The COVID-19 pandemic had a negative effect on step counts among Australians across age groups and genders. However, the effect was relatively small even during the lockdown; in addition, physical activity levels quickly recovered after the lockdown. There was a significant increase in the use of the 10,000 Steps program during the pandemic, which might help minimize the negative effect of the pandemic and confirm the important role of technology-based physical activity programs during times of distress and lockdowns.

Acknowledgments

CV is supported by a Future Leader Fellowship (ID 100427) from the National Heart Foundation of Australia. MJD is supported by a Career Development Fellowship (APP1141606) from the National Health and Medical Research Council. The 10,000 Steps program was funded by the Department of Health (Queensland Health), the Queensland Government (ID 71487).

Conflicts of Interest

None declared.

References

1. 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: US Department of Health and Human Services; 2018. URL: https://health.gov/sites/default/files/2019-09/PAG_Advisory_Committee_Report.pdf [accessed 2021-01-18]
2. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]
3. Hupin D, Roche F, Gremeaux V, Chatard J, Oriol M, Gaspoz J, et al. Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged ≥60 years: A systematic review and meta-analysis. Br J Sports Med 2015 Oct;49(19):1262-1267. [doi: 10.1136/bjsports-2014-094306] [Medline: 26238869]
4. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med 2014 Jan;44(1):81-121 [FREE Full text] [doi: 10.1007/s40279-013-0090-5] [Medline: 24030837]
5. Fitbit Staff. The impact of coronavirus on global activity. Fitbit blog. 2020 Mar 23. URL: https://blog.fitbit.com/covid-19-global-activity/ [accessed 2020-01-18]
6. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide effect of COVID-19 on physical activity: A descriptive study. Ann Intern Med 2020 Nov 03;173(9):767-770. [doi: 10.7326/m20-2665]
7. Ronto P. Exercise is up 88% during COVID-19. RunRepeat. 2020 Jun 18. URL: https://runrepeat.com/exercise-covid-19-study [accessed 2021-01-18]
8. Ding D, Del Pozo Cruz B, Green MA, Bauman AE. Is the COVID-19 lockdown nudging people to be more active: A big data analysis. Br J Sports Med 2020 Oct;54(20):1183-1184. [doi: 10.1136/bjsports-2020-102575] [Medline: 32605932]
9. Capodilupo E, Miller D. Changes in health promoting behavior during COVID-19 physical distancing: Utilizing WHOOP data to examine trends in sleep, activity, and cardiovascular health. medRxiv Preprint posted online on June 9, 2020. [FREE Full text] [doi: 10.1101/2020.06.07.20124685]
10. Stanton R, To QG, Khalesi S, Williams SL, Alley SJ, Thwaite TL, et al. Depression, anxiety and stress during COVID-19: Associations with changes in physical activity, sleep, tobacco and alcohol use in Australian adults. Int J Environ Res Public Health 2020 Jun 07;17(11):4065 [FREE Full text] [doi: 10.3390/ijerph17114065] [Medline: 32517294]
11. Vandelanotte C, Van Itallie A, Brown W, Mummery WK, Duncan MJ. Every step counts: Understanding the success of implementing the 10,000 Steps project. In: Maeder A, Champion S, Moores C, Golley R, editors. Information Technology Based Methods for Health Behaviours: Selected Papers from Global Telehealth 2019. Amsterdam, the Netherlands: IOS Press; Mar 2020:15-30.
12. Chomistek A, Yuan C, Matthews C, Troiano R, Bowles H, Rood J. Physical activity assessment with the ActiGraph GT3X and doubly labeled water. Med Sci Sports Exerc 2017;49:1935. [doi: 10.1249/mss.0000000000001299]
13. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the Carrot Rewards app, a population-level incentive-based intervention promoting step counts across two Canadian provinces: Quasi-experimental study. JMIR Mhealth Uhealth 2018 Sep 20;6(9):e178 [FREE Full text] [doi: 10.2196/mhealth.9912] [Medline: 30148712]
14. Rayward AT, Vandelanotte C, Corry K, Van Itallie A, Duncan MJ. Impact of a social media campaign on reach, uptake, and engagement with a free web- and app-based physical activity intervention: The 10,000 Steps Australia program. Int J Environ Res Public Health 2019 Dec 12;16(24):5076 [FREE Full text] [doi: 10.3390/ijerph16245076] [Medline: 31842383]
15. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: The case of 10,000 Steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
16. Department of Health. Physical distancing for coronavirus (COVID-19). Australian Government. 2020 Sep 23. URL: https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/how-to-protect-yourself-and-others-from-coronavirus-covid-19/physical-distancing-for-coronavirus-covid-19 [accessed 2021-01-18]
17. Snape J. The coronavirus shutdown is here. So what's banned in Australia and what's not? ABC News. 2020 Mar 22. URL: https://www.abc.net.au/news/2020-03-23/coronavirus-shutdown-in-australia-whats-open-whats-closed/12081074 [accessed 2021-01-18]

18. Hall KS, Hyde ET, Bassett DR, Carlson SA, Carnethon MR, Ekelund U, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020 Jun 20;17(1):78 [FREE Full text] [doi: 10.1186/s12966-020-00978-9] [Medline: 32563261]
19. National Health Survey: First results. Australian Bureau of Statistics. 2018 Dec 12. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release [accessed 2021-01-18]
20. Thwaites J. Global report gives Australia an A for coronavirus response but a D on climate. The Conversation. 2020 Jul 05. URL: https://theconversation.com/global-report-gives-australia-an-a-for-coronavirus-response-but-a-d-on-climate-141982 [accessed 2021-01-18]
21. Burton NW, Khan A, Brown WJ. How, where and with whom? Physical activity context preferences of three adult groups at risk of inactivity. Br J Sports Med 2012 Dec;46(16):1125-1131. [doi: 10.1136/bjsports-2011-090554] [Medline: 22267568]
22. Alley SJ, Schoeppe S, Rebar AL, Hayman M, Vandelanotte C. Age differences in physical activity intentions and implementation intention preferences. J Behav Med 2018 Jun;41(3):406-415. [doi: 10.1007/s10865-017-9899-y] [Medline: 29116569]
23. Blanchard-Fields F, Mienaltowski A, Seay RB. Age differences in everyday problem-solving effectiveness: Older adults select more effective strategies for interpersonal problems. J Gerontol B Psychol Sci Soc Sci 2007 Jan;62(1):P61-P64. [doi: 10.1093/geronb/62.1.p61] [Medline: 17284559]
24. John OP, Gross JJ. Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. J Pers 2004 Dec;72(6):1301-1333. [doi: 10.1111/j.1467-6494.2004.00298.x] [Medline: 15509284]
25. Shiota MN, Levenson RW. Effects of aging on experimentally instructed detached reappraisal, positive reappraisal, and emotional behavior suppression. Psychol Aging 2009 Dec;24(4):890-900 [FREE Full text] [doi: 10.1037/a0017896] [Medline: 20025404]
26. Carson V, Adamo K, Rhodes RE. Associations of parenthood with physical activity, sedentary behavior, and sleep. Am J Health Behav 2018 May 01;42(3):80-89. [doi: 10.5993/ajhb.42.3.8]
27. Bellows-Riecken KH, Rhodes RE. A birth of inactivity? A review of physical activity and parenthood. Prev Med 2008 Feb;46(2):99-110. [doi: 10.1016/j.ypmed.2007.08.003] [Medline: 17919713]
28. Egli T, Bland HW, Melton BF, Czech DR. Influence of age, sex, and race on college students' exercise motivation of physical activity. J Am Coll Health 2011;59(5):399-406. [doi: 10.1080/07448481.2010.513074] [Medline: 21500059]
29. Cases: States and territories. Coronavirus (COVID-19) in Australia. URL: https://www.covid19data.com.au/states-and-territories [accessed 2021-01-18]

Edited by G Eysenbach; submitted 28.08.20; peer-reviewed by L van Genugten, H Meggy; comments to author 23.10.20; revised version received 19.11.20; accepted 14.12.20; published 25.01.21.

Please cite as:
To QG, Duncan MJ, Van Itallie A, Vandelanotte C
Impact of COVID-19 on Physical Activity Among 10,000 Steps Members and Engagement With the Program in Australia: Prospective Study
J Med Internet Res 2021;23(1):e23946
URL: http://www.jmir.org/2021/1/e23946/
doi:10.2196/23946
PMID:33449907

Original Paper

Geographic Distribution of Mental Health Problems Among Chinese College Students During the COVID-19 Pandemic: Nationwide, Web-Based Survey Study

Xiaoyan Wu[1,2,3*], PhD, MD; Shuman Tao[2,3,4*], PhD, MD; Yi Zhang[1,2,3*], MD; Shiyue Li[5], MD, PhD; Le Ma[6], MD; Yizhen Yu[7], PhD, MD; Guilong Sun[8], MS; Tingting Li[1,2,3], MPH; Fangbiao Tao[1,2,3], PhD, MD

[1]Department of Maternal, Child and Adolescent Health, School of Public Health, Anhui Medical University, Anhui, China
[2]MOE Key Laboratory of Population Health Across Life Cycle, Anhui, China
[3]NHC Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Anhui, China
[4]Department of Nephrology, The Second Hospital of Anhui Medical University, Anhui, China
[5]School of Health Sciences, Wuhan University, Wuhan, China
[6]School of Public Health, Xi’an Jiaotong University Health Science Center, Xi'an, China
[7]School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, China
[8]South-Central Minzu University, Wuhan, China
[*]these authors contributed equally

Corresponding Author:
Fangbiao Tao, PhD, MD
Department of Maternal, Child and Adolescent Health
School of Public Health
Anhui Medical University
No. 81 Meishan Road
Anhui
China
Phone: 86 055165161169
Email: fbtao@ahmu.edu.cn

Abstract

Background: Since the COVID-19 outbreak was first reported, considerable attention has been drawn to mental health problems among college students.

Objective: We aimed to estimate the prevalence of anxiety and depressive symptoms among college students in different geographical areas of China during the early stage of the COVID-19 outbreak.

Methods: A nationwide cross-sectional survey was conducted among Chinese college students of 16 provinces or municipalities from February 4 to 12, 2020. A web-based survey was adopted to collect information from these college students, including demographics, perceived risk of infection, attitudes toward the epidemic and its control, and mental health status. Anxiety symptoms were assessed using the Generalized Anxiety Disorder scale, and depressive symptoms were assessed using the Patient Health Questionnaire. Chi-square test was used to compare the percentage of perceived risk of infection and attitude toward COVID-19 among college students in different geographic locations. Binary logistic models were used to identify associations between geographic locations and mental health problems after controlling for covariates.

Results: A total of 11,787 participants were analyzed in this study (response rate: 79.7%). The prevalence of anxiety and depressive symptoms among college students was 17.8% (95% CI 17.1%-18.5%) and 25.9% (95% CI 25.1%-26.7%), respectively. After controlling for covariates, current residence area in Wuhan city was found to have a positive association with anxiety symptoms (odds ratio [OR] 1.37, 95% CI 1.11-1.68) and depressive symptoms (OR 1.32, 95% CI 1.09-1.59). Similarly, college location in Wuhan city was found to have a positive association with anxiety symptoms (OR 1.20, 95% CI 1.07-1.35) and depressive symptoms (OR 1.22, 95% CI 1.10-1.36). History of residence in or travel to Wuhan city in the past month was also positively associated with anxiety symptoms (OR 1.62, 95% CI 1.46-1.80) and depressive symptoms (OR 1.48, 95% CI 1.35-1.63). Furthermore, the perceived risk of COVID-19 was higher among students whose college location and current residence area were in Wuhan city, and it was positively associated with anxiety and depressive symptoms.

Conclusions: During the COVID-19 pandemic, mental health problems among Chinese college students were widespread and geographically diverse. Our study results provide further insight for policymakers to develop targeted intervention strategies.

(J Med Internet Res 2021;23(1):e23126) doi:10.2196/23126

KEYWORDS

COVID-19 pandemic; college students; mental health problems; geographic location

Introduction

In December 2019, a new type of coronavirus named SARS-CoV-2 emerged in Wuhan, China [1]. This novel virus was found to cause a pneumonia-like respiratory infection called COVID-19; the disease rapidly spread all over China and turned into a pandemic, affecting most countries globally and putting the entire world on alert [2]. Because this is a novel disease, the preventive and treatment options were not clearly established at the beginning of the outbreak, imposing increased risk for mental health problems associated with COVID-19 [3]. The COVID-19 outbreak has also led to a major education crisis for children and adolescents. It is estimated that approximately 1.5 billion young people (ie, over 90% of all enrolled learners) worldwide were out of education [4], and more than 220 million children and adolescents in China were confined at home as a result of the COVID-19 outbreak [5]. Few studies have examined the effects of COVID-19 on young people, but studies on the general population have reported adverse mental health effects, including increased risks of anxiety, depression, general psychological distress, and posttraumatic stress disorder [6-8]. A recent study in China found that the general public's vicarious traumatization scores were significantly higher than those of front-line nurses [9]. Health professionals and the general public were concerned with the potential adverse effects of the COVID-19 pandemic on the mental health of college students [10]. These concerns are consistent with earlier studies that have reported higher rates of mental health problems among college students following other disease outbreaks, such as SARS [11]. However, not much is known about the mental health effects of large-scale disease outbreaks on college students. Thus far, a few previous studies have allowed us to clarify the factors associated with mental health in a large population of college students after the COVID-19 outbreak.

The Chinese Lunar New Year is the largest annual event of mass travel worldwide [12]. In the context of the COVID-19 outbreak, tracking the migratory flow of college students in China is especially important because almost all of them were in the run-up to the annual Chunyun mass migration since the start of the winter holidays. Spatiotemporal patterns can help detect existing spatially clustered characteristics and high-risk areas, which can lead to a better understanding of the disease epidemic in the spatiotemporal dimension, thus providing reliable information for decision-making of disease prevention and control [13]. A compelling spatial analysis has identified hotspots within and outside of Hubei Province in China [14]. Wuhan is a city with many immigrants, and thousands of people left Wuhan to go to other cities and provinces before the COVID-19 lockdown came into effect. During this time, college students were also headed home for the winter holidays [15].

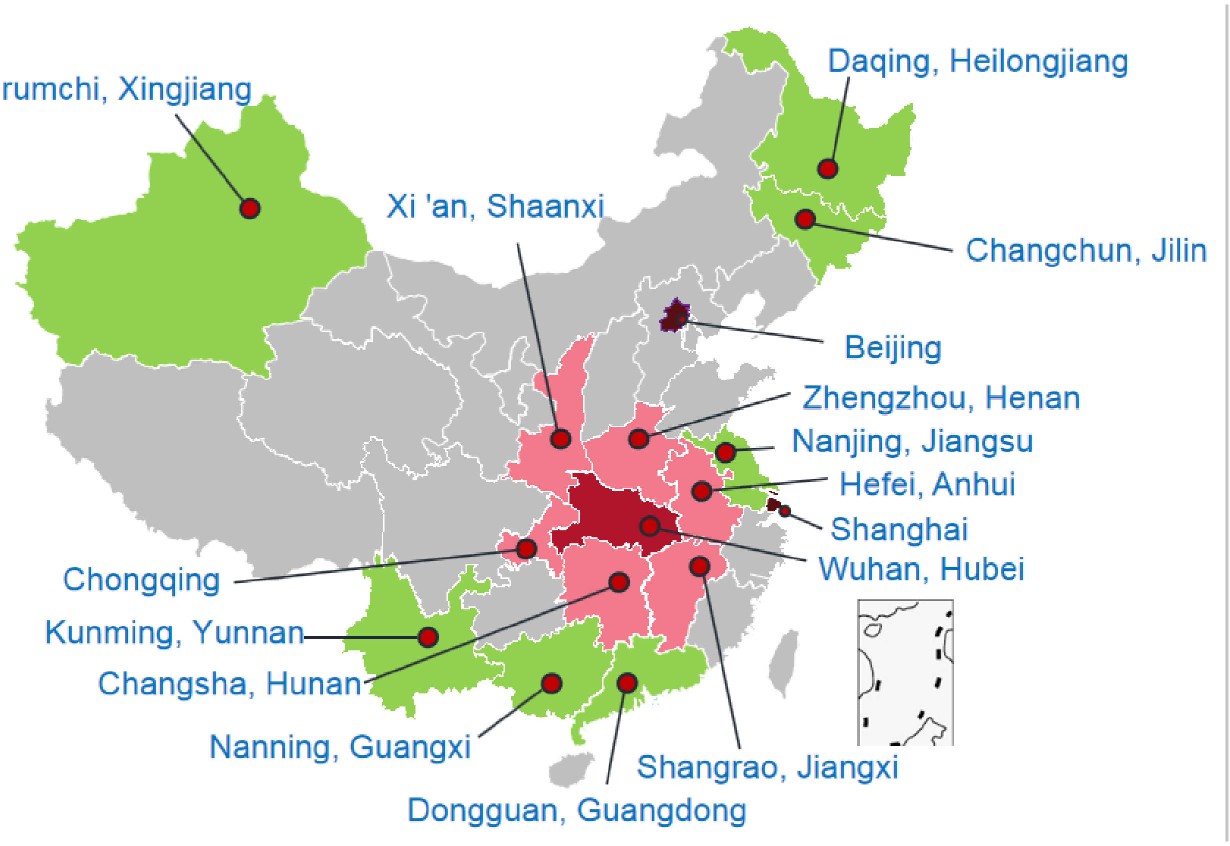
Therefore, we considered this specific population of college students in China to identify target locations for high risk of mental health problems. In this study, we aimed to estimate the prevalence of anxiety and depressive symptoms among college students in different geographical areas of China during the COVID-19 pandemic. It is essential to assess the spatial variability of mental health among Chinese college students during this pandemic, as it could provide useful insights to policymakers for targeted interventions.

Methods

Study Design and Participants

This nationwide, cross-sectional, web-based survey was conducted from February 4 to February 12, 2020. A 2-stage sampling strategy was used. In the first stage, based on the geographic location and cooperation intention, we selected the following 16 provinces or municipalities: Wuhan city, neighboring provinces of Hubei (Henan, Anhui, Jiangxi, Hunan, Chongqing, and Shanxi), first-tier cities (Beijing and Shanghai), and other areas (Jiangsu, Guangdong, Guangxi, Yunnan, Xingjiang, Heilongjiang, and Jilin) (Figure 1). A total of 4 universities in Wuhan, Hubei Province, and 15 universities in other areas or municipalities were randomly selected for the analysis. In the second stage, 100-120 students of each grade (in general, 5 years for medical students and 4 years for nonmedical students) of a faculty were randomly selected from each university; these participants were invited to complete the web-based survey though the Wenjuanxing platform [16]. In total, 14,789 students were selected to participate in this study.

Figure 1. Distribution of study sample based on geographic locations of 16 provinces or municipalities in China.



This study was approved by the Ethics Committee of Anhui Medical University. Electronic informed consent was obtained from all participants before the survey.

Measurements

Geographic Location

Participants were asked to report their college location and current residence area. College locations were divided into the following 4 groups according to the geographic location: Wuhan city (the epicenter of the outbreak), neighboring provinces of Hubei, first-tier cities, and other provinces. Similarly, current residence areas were divided into the following 4 groups according to the geographic location: Wuhan city, other cities in Hubei Province, neighboring provinces of Hubei, and other provinces. Participants were also asked about their history of residence in or travel to Wuhan in the past month via a single question (yes, no).

To further understand the degree to which the COVID-19 pandemic had affected this student population, we divided the participants into the following groups according to the number of confirmed COVID-19 cases in their current residence area: <200, 200-600, 601-1000, and >1000. For this, we referred to the province-wise updates from the World Health Organization’s COVID-19 situation reports [17].

Perceived Risk of Infection and Attitude Toward COVID-19

Perceived risk of infection and attitudes toward COVID-19 were evaluated by the following questions: (1) How likely do you think you are at risk of COVID-19 infection? (2) How likely do you think are your family members at risk of COVID-19 infection? The responses to these questions were assigned the following scores: 1 (much less likely), 2 (less likely), 3 (more likely), and 4 (much more likely). Cronbach α was .876.

Participants’ worries about themselves and community members contracting the infection were evaluated by the following questions: (1) Do you worry about contracting the infection yourself? (2) Do you worry about infection among your community members? The responses were assigned the following scores: 1 (not at all), 2 (only a little), 3 (somewhat worry), and 4 (quite a lot). Cronbach α was .787.

Participants’ attitudes toward the COVID-19 epidemic and its control were evaluated by the following questions: (1) What is your attitude toward the COVID-19 epidemic? The responses to this question were assigned the following scores: 1 (very optimistic), 2 (somehow optimistic), 3 (somehow pessimistic), and 4 (very pessimistic). (2) Do you think the COVID-19 epidemic is hard to control at its current stage? The responses to this question were assigned the following scores: 1 (don’t agree), 2 (don’t agree or disagree), and 3 (agree). Cronbach α was .442.

Mental Health Problems

Participants’ anxiety symptoms were assessed by the Chinese version of the Generalized Anxiety Disorder (GAD-7) scale [18]. Participants were asked to rate the frequency of anxiety symptoms they experienced during the last 2 weeks. The GAD-7 scale contained 7 items that were scored on a 4-point scale, with scores ranging from 0 (not at all), 1 (several days), 2 (more than half the days), and 3 (nearly every day). The total score ranged from 0 to 21, with higher scores indicating higher GAD symptoms. The total score was categorized as mild, moderate, or severe anxiety based on cutoff scores of 5-9, 10-14, and 15-21, respectively.

Depressive symptoms were assessed using the Chinese version of the Patient Health Questionnaire (PHQ-9) [19], which is a 9-item questionnaire designed to screen for depressive symptoms. This questionnaire evaluated the frequency of depressive symptoms experienced by the participants in the last 2 weeks. Participants rated individual items on a 4-point scale, with 0 (not at all), 1 (several days), 2 (more than half the days), and 3 (nearly every day). The total score ranged from 0 to 27, with higher scores indicating higher levels of depression. The total score was categorized as mild, moderate, and severe depressive symptoms based on cutoff scores of 5-9, 10-14, and 15-27, respectively.

Covariates

The following variables were included in the analysis as potential confounders: gender, age, and residence (urban and rural). Participants' physical activity levels were assessed by the question "During the past 7 days, on how many days were you physically active for a total of at least 60 minutes per day? (Add all the time you spend in any kind of physical activity that increased your heart rate and made you breathe hard some of the time.)" [20]. The responses ranged from 0 to 7 days, and a high level of physical activity was defined as at least 3 days of exercise per week. Nighttime sleep duration was assessed by the following questions: "During the past week, when have you usually gone to bed?" "During the past week, when have you usually woken up?" [21]. Screen time was assessed using the following question: "How many hours per day did you spend on the computer (including playing video games or computer games or using a computer for something else) and watching TV/video programs during the past 7 days?" We categorized screen time as <2 hours/day (h/d), 2-4 h/d, and >4 h/d [22-24]. We also examined screen time spent viewing or obtaining COVID-19 information, and categorized it as <0.5 h/d, 0.5-1 h/d, and >1 h/d.

Statistical Analyses

Categorical variables are presented as percentages (%) and frequencies (n). Continuous variables are presented as mean (SD) or median (IQR) values. Chi-square test was used to compare the percentages of perceived risk of infection and attitudes toward the COVID-19 pandemic among college students in different geographic locations. Binary logistic models were used to identify associations between geographic location and mental health problems after controlling for covariates. Variables that were potentially correlated with lifestyle during the COVID-19 pandemic and could cause a bias in the association with mental health, including gender, age, residence, nighttime sleep duration, physical activity, and screen time, were adjusted for in the regression models. We estimated the adjusted odds ratios (ORs) and 95% CIs of independent variables.

Data were analyzed using SPSS (version 23.0; IBM Corp). *P* values <.05 were considered statistically significant.

Results

Characteristics of the Study Sample

Survey responses from a total of 11,787 participants were analyzed in this study (response rate: 79.70%). The mean age of the study participants was 20.45 (SD 1.76) years, and 57.11% (6731/11,787) of all participants were female. Of all participants, 48.02% (5660/11,787) had a rural residence, 5.06% (597/11,787) reported they were currently residing in Wuhan, 41.46% (4887/11,787) attended a college located in Wuhan, and 26.52% (3126/11,787) had a history of residence in or travel to Wuhan in the past month. The prevalence of anxiety and depressive symptoms among the study participants was 17.80% (2098/11,787) and 25.90% (3053/11,787), respectively. Sample characteristics stratified by gender are shown in Table 1. The results of gender differences in risk perception are shown in the table in Multimedia Appendix 1. Our results showed that the high rate of perceived infection risk of participants themselves and their family members was more elevated among female participants than among male participants. Moreover, compared with male participants, female participants were more worried about themselves and their community members contracting the infection. The pessimistic attitude of female participants toward the COVID-19 epidemic was also higher than that observed among male participants. A larger proportion of female participants agreed that COVID-19 is difficult to control at the current stage. Prevalence of anxiety and depressive symptoms among college students with different demographic characteristics is shown in Multimedia Appendices 2 and 3, respectively.

Table 1. Characteristics of study participants by gender.

Characteristic	Total (N=11,787)	Male (n=5056)	Female (n=6731)	*t* test (*df*)	Chi-square (*df*)	*P* value
Age, mean (SD)	20.45 (1.76)	20.44 (1.73)	20.46 (1.78)	-0.39 (11,785)	N/A[a]	.699
Nighttime sleep duration (h/d), mean (SD)	9.57 (1.43)	9.46±1.55	9.65±1.32	-6.99 (11,785)	N/A	<.001
Grade, n (%)				N/A	122.82 (4)	<.001
1	2930 (24.9)	1384 (27.4)	1546 (23)			
2	2609 (22.1)	1201 (23.8)	1408 (20.9)			
3	2667 (22.6)	1153 (22.8)	1514 (22.5)			
4	2314 (19.6)	936 (18.5)	1378 (20.5)			
5	1267 (10.7)	382 (7.6)	885 (13.1)			
Residence, n (%)				N/A	3.63 (1)	.06
Rural	5660 (48)	2479 (49)	3181 (47.3)			
Urban	6127 (52)	2577 (51)	3550 (52.7)			
Current residence area, n (%)				N/A	41.53 (3)	<.001
Wuhan	597 (5.1)	235 (4.6)	362 (5.4)			
Other cities in Hubei	2237 (19)	910 (18)	1327 (19.7)			
Neighboring provinces of Hubei	2750 (23.3)	1323 (26.2)	1427 (21.2)			
Other provinces	6203 (52.6)	2588 (51.2)	3615 (53.7)			
College location, n (%)				N/A	83.13 (3)	<.001
Wuhan	4887 (41.5)	1868 (36.9)	3019 (44.9)			
Neighboring provinces of Hubei	2800 (23.8)	1336 (26.4)	1464 (21.8)			
First-tier cities	900 (7.6)	434 (8.6)	466 (6.9)			
Other provinces	3200 (27.1)	1418 (28)	1782 (26.5)			
History of residence in or travel to Wuhan in the past month, n (%)				N/A	10.51 (1)	.001
Yes	3126 (26.5)	1264 (25)	1862 (27.7)			
No	8661 (73.5)	3792 (75)	4869 (72.3)			
Physical activity, n (%)				N/A	80.08 (1)	<.001
<3d/w	8334 (70.7)	3356 (66.4)	4978 (74)			
≥3d/w	3453 (29.3)	1700 (33.6)	1753 (26)			
Screen time, n (%)				N/A	225.13 (2)	<.001
<2 h/d	2511 (21.3)	1395 (27.6)	1116 (16.6)			
2-4 h/d	3706 (31.4)	1557 (30.8)	2149 (31.9)			
>4 h/d	5570 (47.3)	2104 (41.6)	3466 (51.5)			
Screen time spent viewing or obtaining COVID-19 information, n (%)				N/A	110.80 (2)	<.001
<0.5 h/d	3231 (27.4)	1613 (31.9)	1618 (24)			
0.5-1 h/d	4623 (39.2)	1965 (38.9)	2658 (39.5)			
>1 h/d	3933 (33.4)	1478 (29.2)	2455 (36.5)			
Anxiety symptoms, n (%)				N/A	31.82 (1)	<.001
Yes	2098 (17.8)	784 (15.5)	1314 (19.5)			
No	9689 (82.2)	4272 (84.5)	5417 (80.5)			
Depressive symptoms, n (%)				N/A	57.55 (1)	<.001
Yes	3053 (25.9)	1131 (22.4)	1922 (28.6)			
No	8734 (74.1)	3925 (77.6)	4809 (71.4)			

[a] N/A: not applicable.

Participants whose current residence area and college were located in Wuhan had higher depressive and anxiety symptoms than those from other regions. Moreover, depressive and anxiety symptoms were more elevated in female participants than in male participants. Participants who had a history of residence in or travel to Wuhan in the past month also had higher depressive and anxiety symptoms than those without such history.

Differences Between Current Residence Area or College Location and Perceived Risk of Infection and Attitude Toward COVID-19

Perceived risk of infection and attitudes toward COVID-19 for participants in different current residence areas and college location groups are shown in Multimedia Appendices 4 and 5, respectively. The percentage of high perceived infection risk for individuals was higher among participants who were currently residing in Wuhan and whose college area was located in Wuhan (both *P*<.001). Likewise, the percentage of high perceived infection risk for family members was higher among participants who were currently residing in Wuhan and whose college area was located in Wuhan (both *P*<.001). The percentage of participants who worried very much about themselves and their community members contracting the infection was higher among those whose college and current residence was located in Wuhan (both *P*<.001). The percentage of optimistic attitude toward COVID-19 was the lowest among college students whose current residence and college was located in Wuhan compared to those in the other groups. The highest proportion of college students currently residing in Wuhan agreed with the statement “COVID-19 is hard to control at its current stage” (*P*<.001); nevertheless, a median proportion of college students whose college area was located in Wuhan agreed with this statement (*P*<.001).

Differences Between Perceived Risk of Infection and Attitude Toward COVID-19 and Mental Health

Overall, a strong gradient was observed between different levels of perceived infection risk for individuals themselves and for their family members with regard to the prevalence of anxiety and depression symptoms (Multimedia Appendices 6 and 7). Similarly, a strong gradient was also observed between different levels of worry about infection risk for individuals themselves and for their community members with regard to the prevalence of anxiety and depression symptoms (Multimedia Appendices 8 and 9). Furthermore, a gradient was observed between the different attitudes of college students toward the COVID-19 epidemic and its control and the prevalence of anxiety and depression symptoms (Multimedia Appendices 10 and 11).

Both anxiety and depression symptoms were more prevalent among college students who reported high perceived infection risk for themselves (χ^2_2=220.46 for anxiety, χ^2_2=131.79 for depression; both *P*<.001) and for their family members (χ^2_2=197.52 for anxiety, χ^2_2=136.06 for depression; both *P*<.001). The prevalence of anxiety and depression symptoms was the highest among college students who worried very much about themselves (χ^2_3=285.77 for anxiety, χ^2_3=186.10 for depression; both *P*<.001) and their community members (χ^2_3=178.46 for anxiety, χ^2_3=119.60 for depression; both *P*<.001) contracting the infection. The prevalence of anxiety and depression symptoms was the highest among college students who had a pessimistic attitude toward the COVID-19 epidemic (χ^2_3=444.61 for anxiety, χ^2_3=355.95 for depression; both *P*<.001) and its control (χ^2_2=154.02 for anxiety, χ^2_2=124.84 for depression; both *P*<.001).

Associations Between Geographic Location and Mental Health

Tables 2 and 3 show the estimated ORs for the associations between geographic location and the prevalence of anxiety and depression among Chinese college students. Most associations were evident in both crude and adjusted models.

Table 2. Associations between geographic location and the prevalence of anxiety symptoms among college students in China.

Geographic location	Value, n (%)	Crude odds ratio (95% CI)	Adjusted odds ratio (95% CI)[a]
Current residence area			
Wuhan	136 (22.8)	1.35 (1.10-1.65)*	1.37 (1.11-1.68)*
Other cities in Hubei	368 (16.5)	0.90 (0.79-1.02)	0.92 (0.81-1.06)
Neighboring provinces of Hubei	478 (17.4)	0.96 (0.85-1.08)	0.99 (0.88-1.12)
Other provinces	1116 (18)	1.00	1.00
College location			
Wuhan	967 (19.8)	1.18 (1.05-1.33)*	1.20 (1.07-1.35)*
Neighboring provinces of Hubei	461 (16.5)	0.94 (0.82-1.08)	0.96 (0.84-1.10)
First-tier cities	117 (13)	0.72 (0.58-0.89)*	0.72 (0.58-0.89)*
Other provinces	553 (17.3)	1.00	1.00
History of residence in or travel to Wuhan in the past month			
No	1363 (15.7)	1.00	1.00
Yes	735 (23.5)	1.65 (1.49-1.82)**	1.62 (1.46-1.80)**
Confirmed COVID-19 cases in current residence area			
<200	348 (16.8)	1.00	1.00
200-600	712 (18.2)	1.10 (0.95-1.26)	1.09 (0.95-1.26)
601-1000	513 (17.7)	1.07 (0.92-1.24)	1.10 (0.94-1.28)
>1000	525 (18)	1.09 (0.94-1.26)	1.12 (0.96-1.31)

[a] Controlled for gender, age, residence, grade, nighttime sleep duration, physical activity, and screen time.

*$P<.01$.

**$P<.001$.

Table 3. Associations between geographic location and the prevalence of depressive symptoms among college students in China.

Geographic location	Value, n (%)	Crude odds ratio (95% CI)	Adjusted odds ratio (95% CI)[a]
Current residence area			
Wuhan	187 (31.3)	1.35 (1.12-1.62)*	1.32 (1.09-1.59)*
Other cities in Hubei	579 (25.9)	1.03 (0.92-1.15)	1.02 (0.91-1.15)
Neighboring province of Hubei	718 (26.1)	1.04 (0.94-1.16)	1.08 (0.97-1.20)
Other provinces	1569 (25.3)	1.00	1.00
College location			
Wuhan	1393 (28.5)	1.26 (1.14-1.40)**	1.22 (1.10-1.36)**
Neighboring provinces of Hubei	697 (24.9)	1.05 (0.93-1.18)	1.07 (0.95-1.21)
First-tier cities	195 (21.7)	0.88 (0.73-1.05)	0.89 (0.75-1.07)
Other provinces	768 (24)	1.00	1.00
History of residence in or travel to Wuhan in the past month			
No	2043 (23.6)	1.00	1.00
Yes	1010 (32.3)	1.55 (1.41-1.69)**	1.48 (1.35-1.63)**
Confirmed COVID-19 cases in current residence area			
<200	524 (25.4)	1.00	1.00
200-600	1030 (26.3)	1.05 (0.93-1.87)	1.03 (0.91-1.17)
601-1000	713 (24.7)	0.96 (0.85-1.10)	0.97 (0.85-1.11)
>1000	786 (27)	1.09 (0.96-1.24)	1.06 (0.92-1.21)

[a] Controlled for gender, age, grade, residence, nighttime sleep duration, physical activity, and screen time.

*$P<.01$.

**$P<.001$.

Compared with the college students currently residing in other provinces, those currently residing in Wuhan city were significantly more likely to be at risk for symptoms of anxiety (crude OR 1.35, 95% CI 1.10-1.65, *P*<.01; adjusted OR 1.37, 95% CI 1.11-1.68, *P*<.01) and depression (crude OR 1.35, 95% CI 1.12-1.62, *P*<.01; adjusted OR 1.32, 95% CI 1.09-1.59, *P*<.01). Similarly, compared with the students whose colleges were located in other provinces, those whose colleges were located in Wuhan had a significantly increased risk of anxiety (crude OR=1.18, 95% CI 1.05-1.33, *P*<.01; adjusted OR 1.20, 95% CI 1.07-1.35, *P*<.01) and depression (crude OR 1.26, 95% CI 1.14-1.40, *P*<.001; adjusted OR 1.22, 95% CI 1.10-1.36, *P*<.001). Furthermore, students who had a history of residence in or travel to Wuhan in the past month were significantly more likely to be at risk for symptoms of anxiety and depression than those without such history (anxiety: crude OR 1.65, 95% CI 1.49-1.82, *P*<.001; adjusted OR 1.62, 95% CI 1.46-1.80, *P*<.001; depression: crude OR 1.55, 95% CI 1.41-1.69, *P*<.001; adjusted OR 1.48, 95% CI 1.35-1.63, *P*<.001). All associations were evident in both crude and adjusted models (Tables 2 and 3).

Nevertheless, participants' currently living in areas with a median number (200-600, >1000) of confirmed COVID-19 cases were more likely to be at risk for both anxiety and depression symptoms compared with than were those currently living in areas with <200 confirmed cases; however, these associations were not statistically significant.

Discussion

Principal Findings

The COVID-19 crisis has caused significant public concern in China and around the world. Therefore, there is an urgent need to understand the psychological status of college students in different areas affected by the COVID-19 outbreak. To our knowledge, some studies have described the mental health status of Chinese college students during the COVID-19 epidemic [25,26], but these studies have not evaluated geographical location in detail. In this study, we adopted a web-based questionnaire to explore the association between mental health problems and geographic distribution. Our study provides important findings about the psychological well-being of college students across different geographical locations, which can help policymakers design targeted interventions in order to effectively improve the mental health status of these students.

We found that anxiety and depressive symptoms were positively associated with Wuhan-based geographic locations (ie, current residence area or college location) as well as history of residence in or travel to in Wuhan. Moreover, female students had a higher prevalence of anxiety and depression, which could be because women are considered to be more perceptual, emotional, relatively vulnerable to tension, and have a high incidence of depressive symptoms than men [27,28]. These findings implicate that intervention strategies should primarily focus on college students who studied, lived, or ever traveled to Wuhan city during the COVID-19 pandemic, so as to provide useful insights for policymakers for targeted intervention of COVID-19.

During the outbreak, college students were concerned about the possibility of being infected with COVID-19. In the present study, significant differences were observed between students' current residence area or college location and perceived risk of infection and attitude toward COVID-19. Similarly, de Zwart [29] revealed the relatively high perceived threat for SARS. Public perception of health risk plays a key role in adopting these actions, in people's feelings, and their daily habits [30]. Furthermore, our results revealed that the perceived risk of infection and attitudes toward COVID-19 were both associated with increased risks of anxiety and depressive symptoms. Similar to previous studies, increased worry about contracting infection and lower self-perceived health conditions were significantly associated with higher scores on the Self-Rating Anxiety Scale and Self-Rating Depression Scale [31]. Fear of the unknown increases anxiety levels among healthy individuals as well as those with preexisting mental health conditions [7]. As emotional responses will likely include extreme fear and uncertainty, this can further predict mental health consequences.

During the COVID-19 lockdown period, anxiety and depression among people substantially increased [8]. In the present study, 17.8% and 25.9% of respondents reported anxiety and depressive symptoms, respectively. One of the strengths of our study was that we used nationally representative data collected during the early stage of the COVID-19 outbreak in China and geographic location groups to explore mental health distribution. Currently, a large body of study has utilized web-based or mobile geographic information systems and mapping dashboards for tracking the COVID-19 epidemic [32,33]. The COVID-19 outbreak affected different locations such as Wuhan city or regions outside the Hubei Province, to varying levels [14]. A clear and comprehensive understanding of the effects of the COVID-19 epidemic on the mental health problems of college students residing and attending college across different geographic locations would effectively provide target prevention and control strategies.

We found that anxiety and depressive symptoms were positively associated with the current residence area and college locations based in Wuhan as well as with history of residence or travel to Wuhan. The evidence suggests that, in the early stages of the COVID-19 outbreak in Wuhan, rumors and misinformation were more prevalent in those areas and could cause anxiety and stress about the outbreak [34]. In particular, people who contracted the disease may be more vulnerable to the psychosocial effects than others [3]. Given that college students who were currently residing or attending college in Wuhan, as well as those who had a history of residence in or travel to Wuhan, likely had a higher risk of exposure to infection, there may be greater concern about infection. Moreover, college students residing in Wuhan during the outbreak may have experienced a shortage of personal protective equipment and had inadequate supplies, medical care, and medications during home confinement; all these factors would have likely increased the risks of mental health problems [3].

Our study mapped the effects of geographic location on spatial variability of mental health, and the findings would aid policymakers in developing targeted interventions. Recent studies among university studies during the COVID-19 pandemic have found similar results. For instance, Xin et al [35] reported that the prevalence of depression was 14.8%. Wang et al [36] demonstrated that the prevalence of anxiety and depressive symptoms was 7.7% and 12.2%, respectively. Another study conducted in Hubei Province among medical college students reported that 35.5% of the students experienced depression, and 22.1% experienced anxiety, and most students who were depressed or anxious had mild or moderate conditions [37]. Similar results were found in an American survey, wherein 71% of the participants reported increased stress and anxiety due to the COVID-19 outbreak [38]. The reason for these high levels of depression and anxiety could be attributed to the severity of the epidemic in the United States. These results show that the psychological condition of students during the epidemic cannot be ignored.

Strengths and Limitations

This study evaluated the effects of COVID-19 on mental health among the Chinese youth. The investigation is representative of youth in China during the COVID-19 outbreak, and a relatively high response rate was achieved. The large national sample provided enough statistical power to examine the geographic location and mental health status of the participants. Our results provide further insight into developing targeted intervention strategies, including particular efforts undertaken by vulnerable populations [39,40]. Further research is required to study the effectiveness of prevention and control strategies, such as web-based mental health services [41], to promote mental health among college students during the COVID-19 epidemic.

However, this study has some potential limitations. First, this is a cross-sectional study, so it is difficult to elucidate causal relationships accurately. Additional longitudinal studies are necessary in the future. Nevertheless, we conducted follow-up examinations every 3 months for the study participants, thus enabling further clarification of the causal relationship. Second, despite the convenience of rapid assessment, our web-based survey may have potential respondent bias.

Conclusions

Based on our findings, we can conclude that the geographic location of students' current residence area or college location had a positive association with anxiety and depressive symptoms if based in Wuhan, compared with other areas. Moreover, a history of residence in or travel to Wuhan was also positively associated with anxiety symptoms and depressive symptoms. Thus, based on the geographic location of the college location and current residence area, college students in Wuhan city had a higher risk perception toward COVID-19.

Acknowledgments

The authors would like to thank all participants involved in the study for their inputs. The following researchers are not listed as authors: Heng Meng (Huazhong University of Science & Technology), Hong Yan (Wuhan University), Jinkui Lu and Jianmin Xiang (Shangrao Normal University), Yan Zhang (contacts of Daqing Normal University, Guangxi University, Hohai University), Songli Mei (Jilin University), Asimuguli (Xinjiang Medical University), Xiaoming Lou (Zhengzhou University), Hong Wang (Chongqing Medical University), Lili Pan (Anhui Medical University), Zhaohui Huang (Anhui Medical University), Ying Huang (Kunming Medical University), Lianguo Fu (Bengbu Medical College), Yufeng Wen (Wannan Medical College), Wenjie Gong (Central South University), Jindong Ni (Guangdong Medical University), Yifei Hu (Capital Medical University), Xiaojian Yin (Shanghai Institute of Technology), and Chenwei Pan (Soochow University).

The study was funded by Anhui Medical University Emergency Key Research Project for Novel Coronavirus Pneumonia (YJGG202001).

Authors' Contributions

XW performed data collection and organization and drafted the manuscript. ST and YZ were responsible for contacting cooperative units, participant recruitment, and data collection. SL, LM, YY, GS, and TL were all accountable for participant recruitment and data collection. FT obtained the funding and designed the study. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Gender differences in risk perception.
[DOC File , 48 KB - jmir_v23i1e23126_app1.doc]

Multimedia Appendix 2
Prevalence of anxiety symptom among college students with different demographic characteristics. Values are reported as n(%).
[DOCX File , 22 KB - jmir_v23i1e23126_app2.docx]

Multimedia Appendix 3
Prevalence of depression symptom among college students with different demographic characteristics. Values are reported as n(%).
[DOCX File , 22 KB - jmir_v23i1e23126_app3.docx]

Multimedia Appendix 4
Association between perceived risk of infection, attitude toward COVID-19, and college location. Values reported as n (%).
[DOCX File , 17 KB - jmir_v23i1e23126_app4.docx]

Multimedia Appendix 5
Association between perceived risk of infection, attitude toward COVID-19, and current residence area. Values reported as n (%).
[DOCX File , 18 KB - jmir_v23i1e23126_app5.docx]

Multimedia Appendix 6
Association of perceived infection risk for participants themselves with the prevalence of anxiety and depression symptoms.
[PNG File , 15 KB - jmir_v23i1e23126_app6.png]

Multimedia Appendix 7
Association of perceived infection risk for family members with the prevalence of anxiety and depression symptoms.
[PNG File , 14 KB - jmir_v23i1e23126_app7.png]

Multimedia Appendix 8
Association of worry about infection risk for participants themselves with the prevalence of anxiety and depression symptoms.
[PNG File , 14 KB - jmir_v23i1e23126_app8.png]

Multimedia Appendix 9

Association of worry about infection risk for community members with the prevalence of anxiety and depression symptoms.
[PNG File , 14 KB - jmir_v23i1e23126_app9.png]

Multimedia Appendix 10
Association of the attitude toward the COVID-19 epidemic with the prevalence of anxiety and depression symptoms.
[PNG File , 17 KB - jmir_v23i1e23126_app10.png]

Multimedia Appendix 11
Association of the attitude toward the control of COVID-19 with the prevalence of anxiety and depression symptoms.
[PNG File , 14 KB - jmir_v23i1e23126_app11.png]

References

1. Du Toit A. Outbreak of a novel coronavirus. Nat Rev Microbiol 2020 Mar;18(3):123 [FREE Full text] [doi: 10.1038/s41579-020-0332-0] [Medline: 31988490]
2. Khan S, Ali A, Siddique R, Nabi G. Novel coronavirus is putting the whole world on alert. J Hosp Infect 2020 Mar;104(3):252-253 [FREE Full text] [doi: 10.1016/j.jhin.2020.01.019] [Medline: 32032614]
3. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
4. Lee J. Mental health effects of school closures during COVID-19. Lancet Child Adolesc Health 2020 Jun;4(6):421 [FREE Full text] [doi: 10.1016/S2352-4642(20)30109-7] [Medline: 32302537]
5. Ministry of Education: The national network cloud platform for primary and secondary schools is now available for free (Article in Chinese). Chinanews.com. 2020 Feb 17. URL: http://www.chinanews.com/sh/2020/02-17/9094648.shtml [accessed 2021-01-15]
6. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
7. Usher K, Durkin J, Bhullar N. The COVID-19 pandemic and mental health impacts. Int J Ment Health Nurs 2020 Jun;29(3):315-318 [FREE Full text] [doi: 10.1111/inm.12726] [Medline: 32277578]
8. Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA Intern Med 2020 Jun 01;180(6):817-818. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
9. Li Z, Ge J, Yang M, Feng J, Qiao M, Jiang R, et al. Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control. Brain Behav Immun 2020 Aug;88:916-919 [FREE Full text] [doi: 10.1016/j.bbi.2020.03.007] [Medline: 32169498]
10. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
11. Main A, Zhou Q, Ma Y, Luecken LJ, Liu X. Relations of SARS-related stressors and coping to Chinese college students' psychological adjustment during the 2003 Beijing SARS epidemic. J Couns Psychol 2011 Jul;58(3):410-423. [doi: 10.1037/a0023632] [Medline: 21574694]
12. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? Int J Epidemiol 2020 Jun 01;49(3):717-726 [FREE Full text] [doi: 10.1093/ije/dyaa033] [Medline: 32086938]
13. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 2020 Jun;582(7812):389-394. [doi: 10.1038/s41586-020-2284-y] [Medline: 32349120]
14. Tang W, Liao H, Marley G, Wang Z, Cheng W, Wu D, et al. The changing patterns of coronavirus disease 2019 (COVID-19) in China: a tempogeographic analysis of the severe acute respiratory syndrome coronavirus 2 epidemic. Clin Infect Dis 2020 Jul 28;71(15):818-824 [FREE Full text] [doi: 10.1093/cid/ciaa423] [Medline: 32296826]
15. Chen Z, Zhang Q, Lu Y, Guo Z, Zhang X, Zhang W, et al. Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan, China. Chin Med J (Engl) 2020 May 05;133(9):1044-1050 [FREE Full text] [doi: 10.1097/CM9.0000000000000782] [Medline: 32118644]
16. Wenjuanxing Online Survey Platform. URL: https://www.wjx.cn/ [accessed 2021-01-18]
17. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-01-18]
18. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

19. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 09;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]
20. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Hawkins J, et al. Youth Risk Behavior Surveillance - United States, 2015. MMWR Surveill Summ 2016 Jun 10;65(6):1-174. [doi: 10.15585/mmwr.ss6506a1] [Medline: 27280474]
21. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Research 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4]
22. Sisson SB, Camhi SM, Church TS, Martin CK, Tudor-Locke C, Bouchard C, et al. Leisure time sedentary behavior, occupational/domestic physical activity, and metabolic syndrome in U.S. men and women. Metab Syndr Relat Disord 2009 Dec;7(6):529-536 [FREE Full text] [doi: 10.1089/met.2009.0023] [Medline: 19900152]
23. Sloan RA, Sawada SS, Girdano D, Liu YT, Biddle SJH, Blair SN. Associations of sedentary behavior and physical activity with psychological distress: a cross-sectional study from Singapore. BMC Public Health 2013 Sep 24;13:885 [FREE Full text] [doi: 10.1186/1471-2458-13-885] [Medline: 24063622]
24. Hamer M, Stamatakis E, Mishra GD. Television- and screen-based activity and mental well-being in adults. Am J Prev Med 2010 Apr;38(4):375-380. [doi: 10.1016/j.amepre.2009.12.030] [Medline: 20307805]
25. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
26. Dong L, Bouey J. Public mental health crisis during COVID-19 pandemic, China. Emerg Infect Dis 2020 Jul;26(7):1616-1618 [FREE Full text] [doi: 10.3201/eid2607.200407] [Medline: 32202993]
27. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
28. Sareen J, Erickson J, Medved MI, Asmundson GJG, Enns MW, Stein M, et al. Risk factors for post-injury mental health problems. Depress Anxiety 2013 Apr;30(4):321-327. [doi: 10.1002/da.22077] [Medline: 23408506]
29. de Zwart O, Veldhuijzen IK, Elam G, Aro AR, Abraham T, Bishop GD, et al. Perceived threat, risk perception, and efficacy beliefs related to SARS and other (emerging) infectious diseases: results of an international survey. Int J Behav Med 2009;16(1):30-40 [FREE Full text] [doi: 10.1007/s12529-008-9008-2] [Medline: 19125335]
30. Motta Zanin G, Gentile E, Parisi A, Spasiano D. A preliminary evaluation of the public risk perception related to the COVID-19 health emergency in Italy. Int J Environ Res Public Health 2020 Apr 27;17(9):3024 [FREE Full text] [doi: 10.3390/ijerph17093024] [Medline: 32349253]
31. Lei L, Huang X, Zhang S, Yang J, Yang L, Xu M. Comparison of prevalence and associated factors of anxiety and depression among people affected by versus people unaffected by quarantine during the COVID-19 epidemic in Southwestern China. Med Sci Monit 2020 Apr 20;26:e924609. [doi: 10.12659/msm.924609]
32. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ 2020 Aug 01;728:138884 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138884] [Medline: 32335404]
33. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 2020 Mar 11;19(1):8 [FREE Full text] [doi: 10.1186/s12942-020-00202-8] [Medline: 32160889]
34. Kumar A, Nayar KR. COVID 19 and its mental health consequences. J Ment Health 2020 Apr 27:1-2. [doi: 10.1080/09638237.2020.1757052] [Medline: 32339041]
35. Xin M, Luo S, She R, Yu Y, Li L, Wang S, et al. Negative cognitive and psychological correlates of mandatory quarantine during the initial COVID-19 outbreak in China. Am Psychol 2020;75(5):607-617. [doi: 10.1037/amp0000692] [Medline: 32673008]
36. Wang Z, Yang H, Yang Y, Liu D, Li Z, Zhang X, et al. Prevalence of anxiety and depression symptom, and the demands for psychological knowledge and interventions in college students during COVID-19 epidemic: A large cross-sectional study. J Affect Disord 2020 Oct 01;275:188-193 [FREE Full text] [doi: 10.1016/j.jad.2020.06.034] [Medline: 32734907]
37. Liu J, Zhu Q, Fan W, Makamure J, Zheng C, Wang J. Online mental health survey in a medical college in China during the COVID-19 outbreak. Front Psychiatry 2020;11:459 [FREE Full text] [doi: 10.3389/fpsyt.2020.00459] [Medline: 32574242]
38. Son C, Hegde S, Smith A, Wang X, Sasangohar F. Effects of COVID-19 on college students' mental health in the United States: interview survey study. J Med Internet Res 2020 Sep 03;22(9):e21279 [FREE Full text] [doi: 10.2196/21279] [Medline: 32805704]
39. Shigemura J, Ursano RJ, Morganstein JC, Kurosawa M, Benedek DM. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: Mental health consequences and target populations. Psychiatry Clin Neurosci 2020 Apr;74(4):281-282 [FREE Full text] [doi: 10.1111/pcn.12988] [Medline: 32034840]

40. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
41. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder
OR: odds ratio
PHQ-9: 9-item Patient Health Questionnaire

Edited by R Kukafka; submitted 02.08.20; peer-reviewed by AA Khan, J Li, I Mircheva, A Rovetta; comments to author 11.08.20; revised version received 02.09.20; accepted 10.01.21; published 29.01.21.

Please cite as:
Wu X, Tao S, Zhang Y, Li S, Ma L, Yu Y, Sun G, Li T, Tao F
Geographic Distribution of Mental Health Problems Among Chinese College Students During the COVID-19 Pandemic: Nationwide, Web-Based Survey Study
J Med Internet Res 2021;23(1):e23126
URL: http://www.jmir.org/2021/1/e23126/
doi:10.2196/23126
PMID:33439853

Viewpoint

Digital Health Tools for Managing Noncommunicable Diseases During and After the COVID-19 Pandemic: Perspectives of Patients and Caregivers

Alessandro Monaco[1], MBA, PhD; Katie Palmer[2], PhD; Nicolaj Holm Ravn Faber[3], MA; Irene Kohler[4], MSc; Mitchell Silva[5], PhD; Anita Vatland[6], MSc; Joop van Griensven[7]; Mariano Votta[8], MSc; Donna Walsh[9], BA; Vincent Clay[10], MA; Mehmet Cuneyt Yazicioglu[11], MD; Danute Ducinskiene[12], MD, PhD; Shaantanu Donde[13], MBBS, MSc

[1]École des hautes études commerciales de Paris (HEC Paris), Jouy-en-Josas, France
[2]Oliba, Rome, Italy
[3]The Danish Committee for Health Education, Copenhagen, Denmark
[4]Healthwatch Wiltshire, Trowbridge, United Kingdom
[5]Esperity, Brussels, Belgium
[6]Pårørendealliansen, Oslo, Norway
[7]Pain Alliance Europe, Brussels, Belgium
[8]Cittadinanzattiva/Active Citizenship Network, Rome, Italy
[9]European Federation of Neurological Associations, Brussels, Belgium
[10]Upjohn, Brussels, Belgium
[11]Upjohn, Istanbul, Turkey
[12]Upjohn, Vilnius, Lithuania
[13]Upjohn, Surrey, United Kingdom

Corresponding Author:
Alessandro Monaco, MBA, PhD
École des hautes études commerciales de Paris (HEC Paris)
Rue de la Libération 1
Jouy-en-Josas, 78350
France
Phone: 33 0692948312
Email: alessandro.monaco@hec.edu

Abstract

Background: A reduction in the number of face-to-face medical examinations conducted for patients with noncommunicable diseases (NCDs) during the first wave of the COVID-19 pandemic has led to health care professionals quickly adopting different strategies to communicate with and monitor their patients. Such strategies include the increased use of digital health tools. However, patient preferences, privacy concerns, a lack of regulations, overregulation, and insufficient evidence on the efficacy of digital health tools may have hampered the potential positive benefits of using such tools to manage NCDs.

Objective: This viewpoint aims to discuss the views of an advisory board of patient and caregiver association members. Specifically, we aim to present this advisory board's view on the role of digital health tools in managing patients with NCDs during and after the COVID-19 pandemic, and to identify future directions based on patients' perspectives.

Methods: As an initiative under the NCD Partnership (PARTners in Ncds Engage foR building Strategies to improve Healthy ageing In Patients) model of Upjohn, a web-based advisory board of patient and caregiver advocates was held on July 28, 2020, to bring together key stakeholders from public and private sectors.

Results: The following key themes emerged: (1) technology developers should understand that the goals of patients may differ from those of health care professionals and other stakeholders; (2) patients, health care professionals, caregivers, and other end users need to be involved in the development of digital health tools at the earliest phase possible, to guarantee usability, efficacy, and adoption; (3) digital health tools must be better tailored to people with complex conditions, such as multimorbidity, older age, and cognitive or sensory impairment; and (4) some patients do not want or are unable to use digital health care tools, so adequate alternatives should always be available.

Conclusions: There was consensus that public-private partnership models, such as the Upjohn NCD Partnership, can be effective models that foster innovation by integrating multiple perspectives (eg, patients' perspectives) into the design, development, and implementation of digital and nondigital health tools, with the main overall objective of improving the life of patients with NCDs.

(J Med Internet Res 2021;23(1):e25652) doi:10.2196/25652

KEYWORDS

digital health; information and communication technologies; health technologies; telemedicine; noncommunicable diseases; COVID-19; SARS-CoV-2; patient advocacy; caregivers; ageing; patient empowerment; digital tool; perspective; patient perspective

Introduction

During the COVID-19 pandemic, there has been a justifiable focus on communicable disease prevention and management, which aims to control the risk of infection and transmission. However, noncommunicable diseases (NCDs) are still highly prevalent conditions [1,2], especially in older persons. NCDs accounted for an estimated 71% of worldwide deaths in 2016 [3,4].

During the first wave of the COVID-19 pandemic in 2020, many nonurgent outpatient services were closed to patients with NCDs; appointments were cancelled or postponed to reduce the risk of infection for patients and health care personnel and ease the burden on national health care services. These infection control procedures will have relevant implications for short- and long-term health, disease management, and quality of life of patients with NCDs [5]. This may be especially relevant to patients with complex conditions or multimorbidity (ie, the presence of 2 or more chronic conditions) [6,7]. Such patients usually benefit from integrated care, which includes comprehensive, coordinated care from multidisciplinary experts in multiple sectors [8-10].

Digital health tools include various technologies that can enhance health and health service delivery. Digital health tools can range from instruments that improve communication between health care professionals (HCPs) and patients, such as telemedicine technology, to electronic health records, mobile apps for monitoring symptoms, and medication and appointment reminder systems [11-16]. Digital health tools have been used for a range of purposes during the COVID-19 pandemic, including telemedicine service provision; remote patient monitoring; digital communication between political leaders and scientific authorities; and digital data monitoring for analyzing COVID-19 spread, COVID-19 evolution, and people's perceptions [17,18]. However, most countries lack a regulatory framework to authorize, integrate, and reimburse telemedicine services, and there is large heterogeneity between countries [19].

Even before the COVID-19 pandemic, many have emphasized that digital health tools are important for the prevention and management of NCDs [20], as digital health tools play an essential role in integrated care [9,21], especially in managing multimorbidity [8-10,22], delivering precision medicine [23], and increasing medication adherence [24]. However, older patients are less likely to choose telemedicine visits over office visits than younger patients, and other factors, such as access to technology, also play a role in patients' preferences for service delivery modes [25]. Patients' engagement with digital health tools is affected by multiple factors, including individual motivations, personal life values, approaches for facilitating engagement and recruitment, and the quality of the tool [11].

Since the start of the pandemic, there has been a rapid surge in the use of telemedicine tools and digital health tools for addressing non–COVID-19 medical issues [26]. However, patient preferences, privacy concerns, and insufficient evidence on the efficacy of digital health tools may hamper the potential positive benefits of such tools in the management of NCDs. Since the start of the pandemic, research on the use of digital health tools has mostly focused on how these instruments can be used to screen and diagnose COVID-19 [18]. This raises important questions about how these tools will be used in the health care of patients with NCDs in the postpandemic era. Although evidence-based results on clinical efficacy, ease of use, and cost-effectiveness from the perspective of HCPs, health care institutions, and national health services are important, we must also consider the needs, abilities, and wishes of the patients and caregivers who will ultimately be using and benefiting from these tools. We need to know whether patients want to manage their NCDs with digital health tools in the future, and understand the aspects that should be considered as we move forward to the next stage of the pandemic.

The objective of this viewpoint is to present the perspective of patients with NCDs and caregivers, with regard to the use of digital health tools for the prevention and management of NCDs. This viewpoint does not present a scientific summary of published evidence. Instead, it focuses on the aspects that patients and caregivers deem important for NCD prevention and care, in the context of an advisory board. Our specific aims include (1) describing the evolving scope of digital health tools for NCDs in the advent of the COVID-19 pandemic, and discussing patients' opinions on how these tools can be better designed and developed; (2) outlining ways to increase patient and caregiver engagement to optimize digital health tools; (3) presenting the challenges that patient and caregiver organizations perceive to be barriers to the use of digital health tools; and (4) identifying research priorities for improving the future use of digital health tools in NCD management.

Methods

The content of this viewpoint summarizes the consensus discussion of an advisory board within the NCD Partnership (PARTners in Ncds Engage foR building Strategies to improve Healthy ageing In Patients), which is an initiative that brings together different stakeholders from public and private sectors, so that they can work together to find tools for increasing healthy aging and reducing the burden of NCDs. This initiative involves an institutional collaboration between Upjohn, a division of Pfizer, and the European Innovation Partnership on Active and Healthy Ageing. The NCD Partnership has been engaged in numerous activities, including the creation of several advisory boards that consist of policy makers, physicians, HCPs, researchers, patient and caregiver advocacy groups, patient empowerment organizations, and industry experts.

On July 28, 2020, Upjohn and the European Innovation Partnership on Active and Healthy Ageing organized a web-based advisory board that consisted of patient and caregiver advocacy group members and patient empowerment organizations. The EUPATI (European Patients' Academy on Therapeutic Innovation) Guidance for Patient Involvement in Medicines Research and Development: Health Technology Assessment [27] has suggested that consensus building excercises should be one of the patient involvement activities.

The NCD Partnership has been operating in the field of NCDs and healthy aging for more than 2 years, since 2018. Several advisory boards that consist of researchers, HCPs, and patient advocacy groups have already been created by the NCD Partnership (ie, in September 2018, September 2019, June 2020). Furthermore, the partners of the NCD Partnership have published several papers based on information that was provided in meetings. These papers formed the basis of the advisory board in this viewpoint.

The organizing committee consisted of an expert (AM) in personalized medicine, health care administration, and regulatory affairs; an epidemiologist (KP) with extensive research expertise in NCDs, multimorbidity, and aging disorders; and a physician (SD) with expertise in pharmaceutical medicine, digital initiative leadership, and partnerships for improving patient outcomes. Both AM and KP have extensive experience in the public sector with regard to chairing and conducting consensus meetings, preparing clinical guidelines, and working with researchers, HCPs, and advocacy groups.

In addition to the chairman (AM) and epidemiologist (KP) who led the consensus dialogue, the advisory board included 7 members from patient and caregiver organizations who were recognized by the organizing committee. We aimed to include a wide range of organizations with views on the different aspects of NCDs including caregiver involvement; patient empowerment; health literacy and patient education; patient rights; and specific priority NCD themes, such as aging, pain, and neurology.

In terms of this viewpoint's methodology, we first identified several themes from previous NCD Partnership advisory board meetings that reflected major barriers to NCD management for patients, including integrated care [21], health technology [20], and COVID-19–related changes to the management of risk factors and diseases in patients with NCDs [5]. Afterward, we prepared a draft of talking points based on these themes. This draft was sent to the members of the advisory board. Each member was given time to discuss with their organizations and obtain feedback from patients, caregivers, and other experts from their associations. We then organized individual web-based meetings with the organizing committee and each advisory board member. During these meetings, the advisory board member provided feedback on the talking points. They were asked to add, remove, or modify the discussion points based on comments from their organization, and identify priority areas for discussion. A new set of talking points was then drafted and sent out to advisory board participants before the final meeting. The final meeting was a 1-day, web-based event that lasted for 7 hours. This meeting included several scheduled breaks throughout the day to ensure that any advisory board members with health issues had ample time to rest or move around. AM acted as the chairperson, while the epidemiologist (KP) documented the views, disagreements, and areas of consensus. A qualitative discussion was conducted, and this continued until a consensus was reached. In case a consensus was not reached, a list of areas for future research was made, as reported in the Results section. The meeting was also recorded to allow the chairperson and epidemiologist to review all discussions and identify key themes and recommendations.

Results

Patients' and Caregivers' Perspectives of Digital Health Tools in NCD Management

Summary of Results

The results section of this viewpoint provides a summary of the consensus points that were agreed upon by the patient and caregiver organizations. These consensus points were related to the use of digital health tools for preventing and managing NCDs. We provide several patient and caregiver perspectives and present their opinions on how digital tools can be optimized during and after the COVID-19 pandemic to improve patient outcomes. The results section reflects the perspectives of the patients and caregivers that the advisory board participants represent.

Consensus Theme 1: The Evolving Scope of Digital Health Tools for NCDs in the Advent of the COVID-19 Pandemic, and Patient and Caregiver Perspectives on How These Tools Can Be Better Designed and Developed

The COVID-19 pandemic has prompted a rapid change in the use of digital health tools. The advisory board members felt that the pandemic has provided the opportunity for scaling up the development of these tools so that they can be used to manage both communicable diseases and NCDs. During the pandemic, there was a rise in the use of digital tools for increasing communication between HCPs and patients (eg, telemedicine and video consultation technologies). Digital health tools can be used as an alternative when nonurgent face-to-face consultations are cancelled. Furthermore, these tools can be used for COVID-19 symptom tracking and tracing. These changes may have a positive and negative impact on the future use of digital health tools in health care. During the consensus discussion, the patient and caregiver advocates felt that the pandemic may have triggered the opportunity for citizens to engage in the use of digital tools and participate in studies that collect data on health and NCD prevention. However, some people may be concerned about ethical and privacy issues. They may also be skeptical about replacing face-to-face health care with digital alternatives. There was consensus that digital health development should be part of an overall long-term strategy that advances in parallel with the greater investments that multiple clinical and nonclinical settings have put into medicine and integrated care. The COVID-19 pandemic is likely to result in the reconfiguration of care pathways in the short- and long-term future, and this reconfiguration may lead to a considerable increase in community- and home-based care, wherein digital tools can support patient management. In the following section, we discuss what roles patients and caregivers can play in the design and development of digital health tools, and whether tools used during the COVID-19 pandemic can play a part in the future management of NCDs.

The advisory board members proposed that a comprehensive range of digital health tools that take into account the whole life of an individual (ie, including the epigenetic life cycle) should be used. This includes an individual's preconception, birth, childhood, adulthood, and older age. These tools should also provide methods for primary prevention (eg, control of lifestyle and treatment of risk factors in healthy individuals), secondary prevention, tertiary prevention, and management of NCDs (eg, symptom management for people in a disease state). To prevent NCDs, digital health tools can include exercise monitors, fitness watches, sleep monitors, and apps that monitor lifestyles and NCD risk factors. Tools that focus on disease and symptom management can include a variety of individual-level instruments (eg, medication reminder systems, adherence monitoring systems, etc), organization-level instruments (eg, appointment scheduling systems, electronic health records, etc), or instruments that improve communication between HCPs and patients (eg, teleconsultation technologies, remote symptom monitoring technologies for blood pressure control, wearable glucose monitors, etc). Different individuals who are involved in the care process may be the target users of digital health tools. Such individuals include community nurses, pharmacists, and people who play an important role in patients' care outside of clinical settings [28]. The perspectives of patients and caregivers may differ depending on the goal of the digital health tool and the type of technology used. The advisory board members felt that it was important to identify the target users for each type of technology and have clear objectives for each tool. The profile of patients with NCDs can differ dramatically in terms of their characteristics and the type of disease or diseases that they have. For example, an oncological patient might have very different needs compared to a person with a neurological disease.

There was strong consensus that digital health tools should be used to enhance health care, rather than transition from nondigital NCD management to fully digital NCD management. Although HCPs and patients with NCDs do not usually have the option to manage NCDs through face-to-face consultations during pandemic-related lockdowns and periods of shielding or quarantine, it is important to identify which digital health tools provide the best solutions for patients and HCPs, and ensure that such tools do not replace better alternatives in nonpandemic times. Appropriate digital health tools might differ depending on situational changes, and people need to have a clear motivation for using digital tools instead of nondigital tools. For example, during the COVID-19 pandemic, patients have clearly benefited from teleconsultations because they minimize the contact between patients and HCPs and reduce the risk of contracting SARS-CoV-2. However, many of the advisory board members emphasized that some individuals want face-to-face contact with HCPs once social distancing measures have eased because they miss human contact. This may be relevant to specific patient groups who prefer nondigital contact with HCPs. On the other hand, some patients, such as those with certain psychiatric disorders, may prefer to receive remote medical care in a home setting.

Consensus Theme 2: Methods for Increasing Patient and Caregiver Engagement to Optimize Digital Health Tools

The advisory board members agreed that digital health tools need to demonstrate a clear benefit for patients, to increase their engagement with these tools. If individuals cannot perceive the personal advantages of using a specific tool, it is less likely that they will use it. On the other hand, institutions have analytic tools that can assess the cost-effectiveness of adopting innovative tools and predict the socioeconomic impact of technological solutions. The patient and caregiver advocates emphasized that outcomes need to be meaningful to the patient and take into account individual aspirations. Further, there needs to be clear definitions for what a digital health solution aims to do. For example, a tool can be designed to improve the frequency of communication between HCPs and patients, prevent disease onset, improve adherence to therapies, manage multimorbidity, or help caregivers. During the COVID-19 pandemic, the purpose of digital health tools has mainly been to increase communication between patients and HCPs. However, there are many different goals that can be achieved via digital health tools (eg, symptom monitoring, health tracking via biosensors, risk factor control, and medication adherence), and some of these goals may be important to other end users, such as caregivers, HCPs, or health care systems. Therefore, digital health tool developers should understand that the goals of patients may differ from those of HCPs and other stakeholders. Outcomes that may be deemed more relevant from an HCP perspective (eg, controlling symptoms, increasing medication adherence, etc) might not be as relevant from a patient perspective, in terms of quality of life and other aspects (eg, patients are often more interested in concrete goals, such as visiting grandchildren, being able to go to a concert, gardening, etc). If the HCP perspective is not in line with the end user perspective, any developed digital health tools are likely to be unsuccessful.

The advisory board members felt that HCPs, caregivers, and other end users need to be involved with the development of digital health tools at the earliest phase possible, to increase the chances of developing technologies that people are motivated to use. These technologies should also have a high impact on patient outcomes. Ideally, the involvement of end users should occur before the implementation phase. Carefully designing and developing digital health tools will hopefully improve the life of patients and caregivers by helping them achieve meaningful goals that have an impact on their quality of life.

However, even if better-designed tools become available, this does not guarantee that they will be widely used.

Patients' engagement with digital health tools is affected by multiple factors, including individual motivations and personal life values. However, the advisory board members stressed that during the COVID-19 pandemic, many patients with NCDs have not had many viable alternatives to routine medical care; patients have no choice but to use telemedicine tools. Although telemedicine techniques are essential for helping HCPs maintain contact with patients with NCDs during periods of social distancing and lockdown, there should be a clear discussion and assessment of which digital health tools will be best used in the future management of patients with NCDs, and whether such tools are able to adequately replace or enhance nondigital health care techniques. Research, product development, and training must focus on individuals who will ultimately benefit from digital health tools as well as methods for improving tool usability (ie, improving user interfaces to facilitate adoption).

There was an overall consensus for improving potential users' engagement with digital health tool by developing instruments that focus on the positive aspects of a tool, instead of just the negative aspects (eg, using language such as "prevention and health" rather than "disease and symptoms"). Technologies that monitor changes in symptoms may not provide impactful incentives that adequately engage patients. Such technologies even run the risk of demotivating individuals during periods of NCD progression and symptom deterioration. Thus, digital health tools should have distinct purposes based on target users and their goals. For example, sports watches should be promoted as preventative tools for healthy individuals who want to monitor their fitness and reduce the risk of developing a future NCD, whereas people who already have NCDs might become distressed from receiving continuous feedback on their health data. Therefore, people with NCDs might better engage with tools that focus on more concrete changes that are relevant to their needs. Patients with NCDs are often motivated by specific and personalized goals, such as being able to independently manage their everyday life.

The advisory board members suggested that it is challenging to have individuals engage with digital health tools over long periods of time. People are often more motivated to use health-related technology when they receive a new diagnosis or experience symptoms. However, when their symptoms improve, the urge to use digital health tools can decrease. Thus, in addition to increasing new users' engagement with digital health tools, it is important to find methods for increasing motivation in people who have recovered or have not experienced severe disease symptoms, to prevent reoccurrence.

Increasing patients' engagement with digital health tools can also be achieved by designing or adapting digital health tools to be more user-friendly, and by focusing on the specific limitations (ie, physical, functional, and cognitive limitations) that patients with NCDs may experience. It is imperative to develop strategies that limit the time and effort needed to input data, so that simpler, more time-efficient tools can be created. It is also essential to tailor digital health tools to individuals with limited experience of technology (eg, older persons), and to adapt technologies so that they can be better used by people with specific needs, such as those with cognitive or sensory impairment.

Many people with NCDs, especially those with multimorbidity, are older individuals. The advisory board felt that age is an important factor to consider, because older individuals may be less likely to use digital health tools than younger people, and specific strategies are needed to engage older individuals. For example, isolated older persons who live in remote areas might not have access to technology or may not use the internet. The patient and caregiver advocates agreed that digital health approaches must be sensitive to the needs and wants of older people. Furthermore, the advisory board members questioned whether HCPs should continue to have patients engage with nondigital methods, or try to change the situation by providing patients with tools, technology, or training for using new digital health care tools. There will always be a proportion of the population who will not want or will not be able to use a digital health care solution. Therefore, there should always be adequate alternatives available to those who want them.

Consensus Theme 3: Challenges That Patient and Caregiver Organizations Perceive to Be Barriers to the Use of Digital Health Tools

The advisory board members discussed how to best use digital health tools in the prevention and management of NCDs in the short- and long-term future. When developing innovative solutions, there need to be clear outcomes and benefits for all patients and stakeholders. There was concern that the rapid development of digital health tools might not necessarily result in the best care of patients. Furthermore, whether patients and HCPs like using digital health tools, and whether they prefer digital health tools over nondigital alternatives still needs to be established. Patient advocacy organizations can help to improve knowledge by investigating patients' motivations for using digital health tools. For example, Pain Alliance Europe [29] conducted a web-based survey in 2020 to collect information on the use of eHealth (ie, the use of information and communication technologies for health purposes) and mobile health (mHealth; ie, the use of mobile devices for medical and public health practices) services. Data were collected from 1789 patients with chronic pain from 28 European countries. The responses showed that 46.28% (827/1787) of individuals did not use eHealth or mHealth services. The use of such services varied between different countries and age groups; the frequency of a lack of use was generally higher in older age groups (ie, >50 years) than in younger age groups. One of the main reasons that patients did not use eHealth or mHealth services was because they did not know about them (422/779, 54.17%). Additionally, 16.17% (126/779) of participants did not use such services because they did not see how these services would benefit them as an individual. The most common reason for using eHealth and mHealth services was because patients could use them to self-manage their conditions. With regard to the patients who did use eHealth or mHealth services, almost half (317/472, 48.84%) did not share their health data with anyone, while 9.86% (64/472) of patients shared data with their primary care physician. Future surveys that evaluate patient preferences, motivations, and patterns of use can help identify barriers to the use of digital health tools and different patterns that are based on the types of NCDs and symptoms.

Another major concern of the advisory board members was whether the use of digital health tools in health care will lead to inequalities among patients. Nondigital alternatives are needed for people who are unwilling or unable to use digital health technologies, and specific groups (eg, older persons, migrants, and people with cognitive or sensory impairment) should not be disadvantaged if they do not use digital health tools. Inequalities among patients may be relevant to different aspects of digital health tool use, including access to technology, training, digital health literacy, and motivation to use them. Many individuals have expressed concerns about data privacy and who their information will be shared with (eg, insurance companies, banks, or employers). It is unlikely that patients will fully embrace digital health care tools until privacy concerns are adequately addressed.

Consensus Theme 4: Research Priorities for Improving the Use of Digital Health Tools in Future NCD Management

The advisory board members agreed that before digital health tools are used on a wider scale, high-quality research is needed to investigate how effective they are in the prevention and management of NCDs during the COVID-19 pandemic era and beyond. The patient and caregiver advocates devised several concrete research questions that should be prioritized. The questions about the individual patients' characteristics, needs, and wants, and the questions about how digital health tools will improve patient outcomes and quality of life are particularly important. Specific research objectives are proposed in Textbox 1. The advisory board members were also concerned about the amount of current evidence that is based on clinical outcomes that have been defined and assessed by HCPs, as there is often a lack of patient and caregiver perspectives in the literature. Research should clarify the distinctions between digital health tools that are developed for healthy people (eg, digital health tools that help prevent disease in the long term) and those developed for patients (eg, digital health tools that improve patient outcomes via symptom monitoring or increasing drug adherence). A key topic for research is determining whether motivations for using digital health tools differ among patients with different diseases or characteristics. For example, if the goal is to increase physical activity levels, digital health tool developers need to know which people are driven by face-to-face interventions and which people prefer using wearable devices. We also need to build registries to identify biomarkers and risk factors for so-called "unpreventable diseases." Digital health tools may help with gathering and analyzing data on a wider scale. It is also vital to develop and design health technologies while carefully considering the European Charter of Patients' Rights [30].

Textbox 1. Research questions on digital health tools in noncommunicable disease management, as suggested by patient and caregiver organizations on the advisory board.

Research questions

- How did patients and health care professionals view the use of telemedicine during the first wave of the COVID-19 pandemic? Did these tools have a positive impact on patient management and symptom control in the short and long term?
- How do patients and health care professionals feel about teleconsultations? Do patient outcomes differ depending on whether care is provided through remote telemedicine tools versus face-to-face methods? Are there specific diseases that these solutions work better for (eg, people with social anxiety might prefer teleconsultations)?
- Which type of digital health tools have a greater impact on the outcomes of individuals and is it always a positive impact?
- Do patients and health care professionals want to continue using telemedicine and other digital health tools after the COVID-19 pandemic, and if so, which ones?
- Which demographic groups are more likely to use digital health tools and how can the other groups be encouraged to use tools more frequently?
- How effective is technology in increasing behavioral change? Do changes in behavior affect clinical outcomes or help patients achieve meaningful goals?
- Which clinical changes have an effect on a patient's quality of life and how can health care professionals better assess what patients' individualized goals are?
- Is quality of life improved by better medication adherence or do changes such as symptom control or psychological support have a bigger effect? Do these differ based on disease, age, gender, and other characteristics?
- Can single digital health tools be adapted to the specific goals of individual patients?
- What motivates individual people to use digital health tools? Does motivation differ based on disease type and other patient characteristics?

Discussion

In conclusion, the COVID-19 pandemic has instigated a rapid increase in the use of digital health tools, and has highlighted the need for developing better tools that are adequately suited to the future needs of both HCPs and patients. It is still unclear how the COVID-19 pandemic will develop; it seems likely that we will face new outbreaks and repeated lockdowns that will result in a renewed disruption of NCD health services and management. Digital health tools can help with providing alternatives for nonurgent, face-to-face medical and care services during lockdown periods. However, digital health tools can also be used to further enhance NCD prevention and management by integrating them into health care services during nonpandemic times. Furthermore, patients' and caregivers' opinions are essential for driving the development of effective tools with the maximum potential impact for end users. Patient and caregiver advocacy groups, citizen networks, and empowerment organizations need to be better integrated into all phases of digital health tool development. This will ensure that meaningful patient and public involvement occurs. Concrete goals need to be set for increasing patients', caregivers', and HCPs' digital literacy. We also need to build and enhance well-connected infrastructures that allow for equitable access to digital health tools. Current regulatory and reimbursement frameworks also need to be reviewed so that they can be adapted to the changing health care pathways that are likely to be increasingly supported by future digital health tools.

We need to clarify how to facilitate patient and caregiver association involvement in the research, development, design, and innovation of care processes and digital health care tools. Our viewpoint highlights the importance of public-private partnership models (eg, the NCD Partnership) and other initiatives (eg, the EUPATI [27], the Partners Forum [31], and the NCD Alliance [32]) in fostering innovation to achieve meaningful goals. The NCD Partnership is an initiative that aims to create a platform for debate and facilitate the collaboration of policy makers (eg, the European Commission), HCPs, industry experts, researchers, patients, and caregivers, by creating advisory boards, implementing collaborative research projects, and conducting other activities. It has been emphasized that successful plans for addressing NCDs require international-level, national-level, and local-level partnerships among national governments, private sectors, and civil societies, and the involvement of a wide range of key actors, including patients and their families [33]. Other multistakeholder partnerships, such as the EUPATI, have emphasized the importance of patient involvement in health technology assessment [27], and have suggested that consensus-building exercises can be used to improve patient involvement. Strengthening the collaboration between public and private organizations that have a strong involvement in civic and patient associations (ie, in terms of health literacy and digital health literacy) can help increase people's ability to obtain, read, understand, and use information that relates to digital technologies. By bringing together multiple stakeholders, including patients, into the design, development, and implementation of digital health tools, we can make advancements that will help align the goals of all partners who share an overall objective of improving the life of patients with NCDs.

Conflicts of Interest

VC is employed by Pfizer Upjohn and DD, MCY, and SD are former employees of Pfizer Upjohn. VC, DD, MCY, and SD are also Pfizer stockholders. AM, KP, NHRF, IK, MS, AV, JVG, MV, and DW have no conflicts of interest to declare.

References

1. Population projections - Population (Demography, Migration and Projections). Eurostat. URL: https://ec.europa.eu/eurostat/web/population-demography-migration-projections/population-projections-data [accessed 2021-01-20]
2. Eurostat. URL: https://ec.europa.eu/eurostat/data/database [accessed 2020-05-03]
3. NCD Countdown 2030 collaborators. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. Lancet 2018 Sep 22;392(10152):1072-1088. [doi: 10.1016/S0140-6736(18)31992-5] [Medline: 30264707]
4. Disease burden and mortality estimates. World Health Organization. URL: http://www.who.int/healthinfo/global_burden_disease/estimates/en/ [accessed 2020-09-16]
5. Palmer K, Monaco A, Kivipelto M, Onder G, Maggi S, Michel JP, et al. The potential long-term impact of the COVID-19 outbreak on patients with non-communicable diseases in Europe: consequences for healthy ageing. Aging Clin Exp Res 2020 Jul;32(7):1189-1194 [FREE Full text] [doi: 10.1007/s40520-020-01601-4] [Medline: 32458356]
6. Xu X, Mishra GD, Jones M. Evidence on multimorbidity from definition to intervention: An overview of systematic reviews. Ageing Res Rev 2017 Aug;37:53-68. [doi: 10.1016/j.arr.2017.05.003] [Medline: 28511964]
7. Beard JR, Officer A, de Carvalho IA, Sadana R, Pot AM, Michel JP, et al. The World report on ageing and health: a policy framework for healthy ageing. Lancet 2016 May 21;387(10033):2145-2154 [FREE Full text] [doi: 10.1016/S0140-6736(15)00516-4] [Medline: 26520231]
8. Hopman P, de Bruin SR, Forjaz MJ, Rodriguez-Blazquez C, Tonnara G, Lemmens LC, et al. Effectiveness of comprehensive care programs for patients with multiple chronic conditions or frailty: A systematic literature review. Health Policy 2016 Jul;120(7):818-832. [doi: 10.1016/j.healthpol.2016.04.002] [Medline: 27114104]
9. Palmer K, Marengoni A, Forjaz MJ, Jureviciene E, Laatikainen T, Mammarella F, Joint Action on Chronic Diseases and Promoting Healthy Ageing Across the Life Cycle (JA-CHRODIS). Multimorbidity care model: Recommendations from the consensus meeting of the Joint Action on Chronic Diseases and Promoting Healthy Ageing across the Life Cycle (JA-CHRODIS). Health Policy 2018 Jan;122(1):4-11. [doi: 10.1016/j.healthpol.2017.09.006] [Medline: 28967492]
10. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001;20(6):64-78. [doi: 10.1377/hlthaff.20.6.64] [Medline: 11816692]

11. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
12. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
13. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016 Feb;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
14. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
15. Posadzki P, Mastellos N, Ryan R, Gunn LH, Felix LM, Pappas Y, et al. Automated telephone communication systems for preventive healthcare and management of long-term conditions. Cochrane Database Syst Rev 2016 Dec 14;12:CD009921 [FREE Full text] [doi: 10.1002/14651858.CD009921.pub2] [Medline: 27960229]
16. Robotham D, Satkunanathan S, Reynolds J, Stahl D, Wykes T. Using digital notifications to improve attendance in clinic: systematic review and meta-analysis. BMJ Open 2016 Oct 24;6(10):e012116 [FREE Full text] [doi: 10.1136/bmjopen-2016-012116] [Medline: 27798006]
17. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
18. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
19. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
20. Monaco A, Maggi S, De Cola P, Hassan TA, Palmer K, Donde S. Information and communication technology for increasing healthy ageing in people with non-communicable diseases: identifying challenges and further areas for development. Aging Clin Exp Res 2019 Nov;31(11):1689-1693 [FREE Full text] [doi: 10.1007/s40520-019-01258-8] [Medline: 31317518]
21. Monaco A, Palmer K, Marengoni A, Maggi S, Hassan TA, Donde S. Integrated care for the management of ageing-related non-communicable diseases: current gaps and future directions. Aging Clin Exp Res 2020 Jul;32(7):1353-1358 [FREE Full text] [doi: 10.1007/s40520-020-01533-z] [Medline: 32277438]
22. Rodriguez-Blazquez C, Forjaz MJ, Gimeno-Miguel A, Bliek-Bueno K, Poblador-Plou B, Luengo-Broto SP, et al. Assessing the Pilot Implementation of the Integrated Multimorbidity Care Model in Five European Settings: Results from the Joint Action CHRODIS-PLUS. Int J Environ Res Public Health 2020 Jul 22;17(15):5268 [FREE Full text] [doi: 10.3390/ijerph17155268] [Medline: 32707791]
23. Onder G, Bernabei R, Vetrano DL, Palmer K, Marengoni A. Facing multimorbidity in the precision medicine era. Mech Ageing Dev 2020 Sep;190:111287. [doi: 10.1016/j.mad.2020.111287] [Medline: 32562614]
24. Marengoni A, Monaco A, Costa E, Cherubini A, Prados-Torres A, Muth C, et al. Strategies to Improve Medication Adherence in Older Persons: Consensus Statement from the Senior Italia Federanziani Advisory Board. Drugs Aging 2016 Sep;33(9):629-637. [doi: 10.1007/s40266-016-0387-9] [Medline: 27444182]
25. Reed ME, Huang J, Graetz I, Lee C, Muelly E, Kennedy C, et al. Patient Characteristics Associated With Choosing a Telemedicine Visit vs Office Visit With the Same Primary Care Clinicians. JAMA Netw Open 2020 Jun 01;3(6):e205873 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5873] [Medline: 32585018]
26. Anthony Jnr B. Use of Telemedicine and Virtual Care for Remote Treatment in Response to COVID-19 Pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
27. Hunter A, Facey K, Thomas V, Haerry D, Warner K, Klingmann I, et al. EUPATI Guidance for Patient Involvement in Medicines Research and Development: Health Technology Assessment. Front Med (Lausanne) 2018;5:231 [FREE Full text] [doi: 10.3389/fmed.2018.00231] [Medline: 30238004]
28. The key role played by pharmacies, according to the citizens. Health Innovation | Innovation News Network. 2020 Jun 04. URL: https://www.innovationnewsnetwork.com/the-key-role-played-by-pharmacies-according-to-the-citizens/5370 [accessed 2020-10-20]
29. Survey Long Report. Pain Alliance Europe. URL: https://pae-eu.eu/wp-content/uploads/2020/10/PAE-2020-Survey-long_report_final.pdf [accessed 2020-10-20]
30. European Charter of Patients' Rights. European Commission. 2002 Nov. URL: https://ec.europa.eu/health/ph_overview/co_operation/mobility/docs/health_services_co108_en.pdf [accessed 2020-10-20]

31. Hospedales CJ, Jané-Llopis E. A multistakeholder platform to promote health and prevent noncommunicable diseases in the region of the Americas: the Pan American Health Organization partners forum for action. J Health Commun 2011 Aug;16 Suppl 2:191-200. [doi: 10.1080/10810730.2011.601245] [Medline: 21916722]
32. Dain K. A new chapter for the NCD Alliance: stronger together. Lancet 2017 May 27;389(10084):2089-2090. [doi: 10.1016/S0140-6736(17)31372-7] [Medline: 28560994]
33. Afshari M, Teymourlouy AA, Asadi-Lari M, Maleki M. Global Health diplomacy for noncommunicable diseases prevention and control: a systematic review. Global Health 2020 May 06;16(1):41 [FREE Full text] [doi: 10.1186/s12992-020-00572-5] [Medline: 32375804]

Abbreviations

EUPATI: European Patients' Academy on Therapeutic Innovation
HCP: health care professional
mHealth: mobile health
NCD: noncommunicable diseases
NCD Partnership: PARTners in Ncds Engage foR building Strategies to improve Healthy ageing In Patients

---

Edited by G Eysenbach; submitted 13.11.20; peer-reviewed by G Aguayo, J Taylor; comments to author 04.12.20; revised version received 24.12.20; accepted 13.01.21; published 29.01.21.

Please cite as:
Monaco A, Palmer K, Holm Ravn Faber N, Kohler I, Silva M, Vatland A, van Griensven J, Votta M, Walsh D, Clay V, Yazicioglu MC, Ducinskiene D, Donde S
Digital Health Tools for Managing Noncommunicable Diseases During and After the COVID-19 Pandemic: Perspectives of Patients and Caregivers
J Med Internet Res 2021;23(1):e25652
URL: http://www.jmir.org/2021/1/e25652/
doi:10.2196/25652
PMID:33464206

Original Paper

Application of In-Home Monitoring Data to Transition Decisions in Continuing Care Retirement Communities: Usability Study

Katherine Wild[1], PhD; Nicole Sharma[1], BA; Nora Mattek[1], MPH; Jason Karlawish[2,3,4], MD; Thomas Riley[1], BA; Jeffrey Kaye[1,5], MD

[1]Department of Neurology, Oregon Center for Aging and Technology, Oregon Health and Science University, Portland, OR, United States

[2]Department of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Neurology, University of Pennsylvania, Philadelphia, PA, United States

[5]Department of Biomedical Engineering, Oregon Center for Aging and Technology, Oregon Health and Science University, Portland, OR, United States

Corresponding Author:
Katherine Wild, PhD
Department of Neurology
Oregon Center for Aging and Technology
Oregon Health and Science University
Portland, OR
United States
Phone: 1 503 494 6975
Email: wildk@ohsu.edu

Abstract

Background: Continuous in-home monitoring of older adults can provide rich and sensitive data capturing subtle behavioral and cognitive changes. Our previous work has identified multiple metrics that describe meaningful trends in daily activities over time. The continuous, multidomain nature of this technology may also serve to inform caregivers of the need for higher levels of care to maintain the health and safety of at-risk older adults. Accordingly, care decisions can be based on objective, systematically assessed real-time data.

Objective: This study deployed a suite of in-home monitoring technologies to detect changing levels of care needs in residents of independent living units in 7 retirement communities and to assess the efficacy of computer-based tools in informing decisions regarding care transitions.

Methods: Continuous activity data were presented via an interactive, web-based tool to the staff identified in each facility who were involved in decisions regarding transitions in care among residents. Comparisons were planned between outcomes for residents whose data were shared and those whose data were not made available to the staff. Staff use of the data dashboard was monitored throughout the study, and exit interviews with the staff were conducted to explicate staff interaction with the data platform. Residents were sent weekly self-report questionnaires to document any health- or care-related changes.

Results: During the study period, 30 of the 95 residents (32%) reported at least one incidence of new or increased provision of care; 6 residents made a permanent move to a higher level of care within their communities. Despite initial enthusiasm and an iterative process of refinement of measures and modes of data presentation based on staff input, actual inspection and therefore the use of resident data were well below expectation. In total, 11 of the 25 staff participants (44%) logged in to the activity dashboard throughout the study. Survey data and in-depth interviews provided insight into the mismatch between intended and actual use.

Conclusions: Most continuous in-home monitoring technology acceptance models focus on perceived usefulness and ease of use and equate the intent to use technology with actual use. Our experience suggests otherwise. We found that multiple intervening variables exist between perceived usefulness, intent to use, and actual use. Ethical, institutional, and social factors are considered in their roles as determinants of use.

(J Med Internet Res 2021;23(1):e18806) doi:10.2196/18806

KEYWORDS

technology; remote sensing technology; care transition

Introduction

Although factors associated with older adults' moves to residential facilities have been well described [1-4], our understanding of the reasons for changes in the level of care in persons who live in continuing care retirement communities (CCRCs) is limited. However, there appears to be some overlap with transitions from independent community living. Increased confusion, loss of mobility, medication nonadherence, and reduced socialization have been identified as predictors of movement to higher levels of care in a CCRC [5-8].

A delay in identification of increased dependence or inability to perform routine self-care activities can result in costly and potentially dangerous outcomes for at-risk residents. Typically, decisions about care needs and transitions in the levels of care for older adults living in CCRCs rely on communication and coordination among professional staff members. Ideally, these decisions are based on the evaluations of relevant health and behavioral changes. As Couture et al [9] have argued, the decision-making process is best accomplished by input from all stakeholders across health care professions as well as from family members and the residents themselves. Georgiou et al [10] identified barriers to optimal communication in residential care facilities, which may have a negative impact on the provision of quality care related to efficient and timely transitions to different levels of care. Another challenge to judicious care transitions is the relative inability to detect changes in a resident's care needs before an acute event occurs.

It has been noted that policies and procedures may not be suitable for efficient and timely information transfer. Kelsey et al [5] reported that policies for transfer from one level of care to another vary across facilities, ranging from a multidisciplinary team approach to decision making by a facility manager or administrator. They recommend that future research attention be paid to the ultimate appropriateness of resident transfers to higher levels of care. In the long term, codification of procedures and identification of factors contributing to moves within a residential community can serve to reduce resistance and misunderstanding and may have the potential to enhance safety and prolong independence in at-risk older adults.

A lack of formalized protocols or objective behavioral markers to guide the process may contribute to the uncertainty and divergent views associated with the determination of residents' needs for transition from independent to more assisted living within a continuing care setting. One way to compile objective, systematically assessed activity data is through continuous monitoring technologies. Demiris and Thompson [11] have cited the value of in-home activity monitoring technologies in delivering large, individually anchored data sets that are *useful, meaningful, and actionable*.

Systems that improve our ability to unobtrusively monitor important health changes because of chronic disease and aging could allow timely intervention that prevents avoidable loss of independence. Continuous collection of health and activity information in the home can enable early identification of clinically significant changes. Our experience in examining the feasibility of deploying a comprehensive, ubiquitous sensing platform in the homes of older adults has been reported previously [12,13]. ORCATECH (Oregon Center for Aging & Technology) research has demonstrated the sensitivity of unobtrusive in-home technology to detect early changes in medication management capacity [14,15], patterns of mobility [16-18], nighttime sleep behaviors [19], computer use [20,21], and driving [22]. Such technologies can provide important information in guiding decisions regarding increased care needs. We aimed to test the hypothesis that providing objective and continuous data from home-based technologies to the care teams in retirement communities will result in fewer transitions to higher levels of care through early identification of behavioral or activity changes that lead to increased in-home assistance.

In this paper, we report on the results of Ambient Independence Measures for Guiding Care Transitions (AIMS), a study that provided designated staff at 7 continuing care residential communities with access to an automated, continuous data monitoring platform via a web-based dashboard that collected residents' behavioral and physiological sensor-based independence metrics. Owing to low use of data by the staff, the trial was not able to adequately evaluate the primary hypothesis that providing these data will result in fewer transitions to higher levels of care and increase in-home assistance because of early identification of potential problems. We describe the procedures implemented to maintain staff engagement and consider the challenges of new technology adoption in residential facilities. Using exit interviews with the staff, we examined the reasons for the low use of the data to recommend how studies of home monitoring of health and activity can be improved.

Methods

Overview

Before initiating the trial, focus group sessions were conducted with care transition teams at the participating retirement communities to better understand their routine process of making decisions about when residents need to transition to a higher level of care and to receive feedback on the proposed AIMS data provision interface. Their recommendations were incorporated into the final dashboard interface where feasible. Residents were then monitored for 3 years, with behavior and activity data supplied to identified staff for half the study residents.

Recruitment

Older Adult Participants

Individuals were recruited from existing registries of volunteers for research residing in one of 7 retirement communities in the Portland metropolitan area and during presentations given at these communities about the project. These residential facilities participated in previous and ongoing research about technology and aging. Demographic data were collected at the time of enrollment in this study or other ORCATECH studies. Participants were aged 75 years or older, independently living in an apartment that was larger than one room, living alone, not demented, and of average health for their age (ie, without a medical illness that would limit physical participation or possibly lead to death over the next 36 months). As part of their participation in this study, the residents agreed to have their in-home activity data shared with their facility staff and were required to have internet service and to be computer users. The cohort was allocated to having the staff view their in-home activity via the web-based dashboard versus no viewing, using simple randomization (computer-generated assignment) at a 1:1 ratio, with an enrollment goal of 50 residents per group. The observation period for the study was 3 years.

Resident participants were instructed to live their lives without any specific health or activity intervention. They replied once a week to an email that directly queried them regarding health or activity change (mood, pain, loneliness, falls, hospital visits, visitors, and limited activity due to health) as well as the need for new or additional care.

Staff Participants

A total of 25 staff members at these facilities were self-identified as participants in decision making around residents' transitions in levels of care. They were recruited to this study during regularly scheduled staff meetings. Staff were aged between 21 and 66 years (mean 39.6, SD 10.4 years) and were employed at their present facility for an average of 10 years (range 2 weeks to 16 years). Staff levels of education ranged from high school graduate to master's degree; a majority (17/25, 68%) had obtained at least a college degree. Their job titles varied and included directors, nurses, social workers, and resident care coordinators.

Monitoring Platform

The details of the home technology system have been described previously [12,13]. In brief, sensors and other in-home technologies are deployed to continuously monitor daily activities. In response to specific queries from the research team regarding the dashboard content and interface, the staff made multiple concrete suggestions on all aspects, ranging from how to better navigate the site to changing the symbols used to represent alerts for changes in activity status. On the basis of the input from the care teams, a study protocol was finalized, which focused on home-based sensors and devices whose data were perceived as indicating that major functions had changed and thus could influence independent living decisions. The selected metrics included mobility, physiological health, nighttime behaviors, medication adherence, socialization, cognitive function, and self-reported health changes via a web-based weekly health form (Table 1).

Table 1. Ambient Independence Measures for Guiding Care Transitions study metrics and devices.

Core functions	Measures	Sensors or devices used
Physical capacity and mobility	Total daily activity, number of room transitions, typing speed, time out of home	PIR[a] motion sensors and contact sensors, computer use metrics (eg, keyboard trigrams)
Walking speed	Median weekly walking speed from multiple daily walks	PIR motion sensor line
Sleep and nighttime behavior	Time of awakening in the morning, time spent in bed at night, wake after sleep onset, times up at night, sleep latency	PIR motion sensors
Physiologic health	Daily BMI, morning pulse	Bioimpedance scale
Medication adherence	Percentage of doses missed in a 7-day period relative to prescribed schedule	MedTracker electronic pillbox
Socialization and engagement	Time out of home, time alone, phone call patterns, online computer activity (email and social networking sites)	PIR motion sensors, contact sensors, personal computer, phone monitors
Cognitive function	Time to complete online tasks, mouse movements, prospective memory for medication	Personal computer or tablet, MedTracker
Health and life events	Online self-report: emergency room, doctor, hospital visits, home visitors, mood, pain, loneliness, falls, injuries, change in home space, home assistance received, change in medications	Personal computer or tablet

[a] PIR: passive infrared.

Resident homes were installed with a sensor platform consisting of passive infrared motion and contact sensors, MedTrackers, weight scales, and software to capture all metrics as well as residents' computer use. A web-based reporting tool was developed to track these data, identify outlying data and trends, and provide staff with access to views of these data on a variety of timescales in a dashboard format that was available to the participating facilities' care transition teams. Figure 1 presents an example of a dashboard data summary for a single resident across various metrics over time. The dashboard interface was designed to be interactive, allowing for the manipulation of time scale, activity, and residents of interest. It also provided alerts to the staff for behaviors and activities that fell outside the range of normal for any particular resident. The basic approach for alerts was to develop a baseline model of typical measures for each individual over multiple weeks and monitor these measures on a regular basis (eg, weekly) for trends away from the norm. Alerts were embedded in data summaries for individual behaviors and activities. Figure 2 depicts a sample resident's summary graphic, where outlying behaviors are represented by an alert (!) and green checkmarks indicate the usual behavior *for that resident*.

Figure 1. Sample dashboard display (screenshot) of continuously assessed in-home activity metrics in Ambient Independence Measures for Guiding Care Transitions (AIMS) residents. The display, showing data at the individual level, can be customized by the user to show higher-level summaries, single metrics, numerical detail, and different windows of time. In this custom view, multiple metrics are displayed (eg, time out of home, physiologic measures, sleep measures, bathroom trips). The gray shading indicates preset ranges.

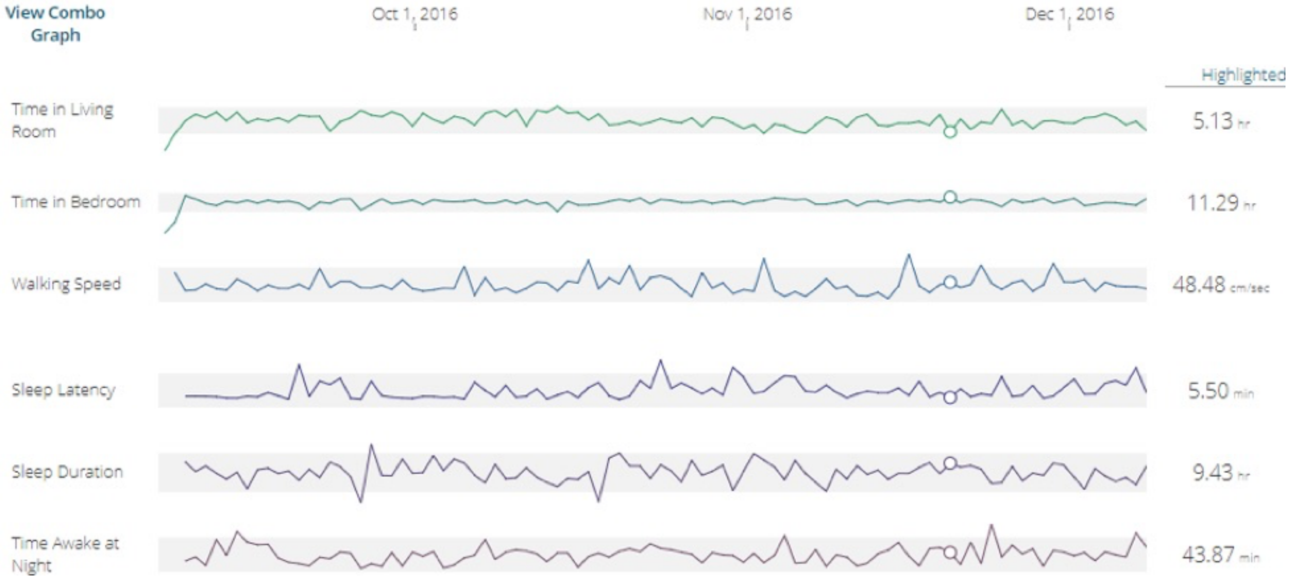


Figure 2. Sample dashboard display of residents' summary data. Exclamation points indicate departure from the usual level of behavior or activity for each resident (Temperature and Carbon Dioxide refer to environmental metrics and were not included in this study).

	Weight	Walking Speed	Medication Adherence	Phone Duration	Temperature	Carbon Dioxide	Room Transitions	Gross Activity	Self Report
Subject 339	✓	✓	✓	✓	✓	✓	✓	✓	✓
Subject 1226	✓	✓	✓	!	✓	✓	✓	!	✓
Subject 1138	!	!	✓	✓	✓	✓	!	✓	✓
Subject 1222	✓	✓	✓	!	✓	!	!	!	✓
Subject 1127	✓	!	✓	✓	!	✓	✓	!	✓
Subject 1149	✓	!	✓	✓	✓	✓	✓	✓	✓
Subject 880	!	✓	✓	✓	✓	✓	✓	✓	✓

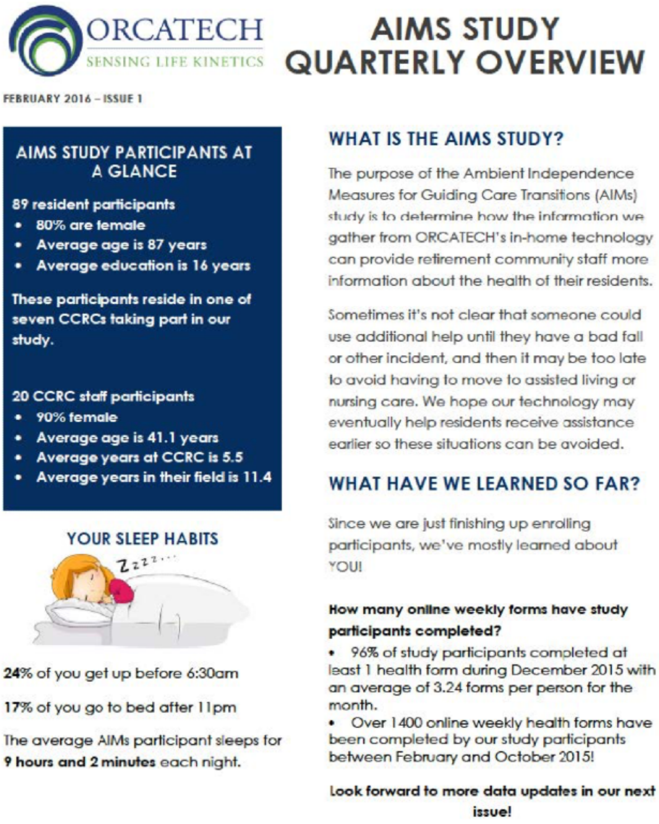
Staff Training and Engagement

Facility staff who had been identified as part of the transition assessment team and were willing to participate in the study were scheduled for 1-hour training sessions by AIMS study staff. Each staff member also received a printed dashboard user guide with detailed instructions on the use of the dashboard and its functionalities in terms of resident activity categories, data summaries, and useful comparisons across activities and time frames. Contact information for additional help was included, and technicians returned to sites to provide additional training as needed. During the study period, minimal additional training was requested; our technician made 2 additional visits to facilities and provided occasional help by telephone. These data are not tabulated.

To maintain engagement with the project, we mailed quarterly newsletters to residents and staff participants (Figure 3).

Newsletters contained study updates regarding recruitment and participation, illustrative dashboard screenshots and data summaries, and *fun facts* based on data collected. For example, one newsletter presented data summarizing the sleep habits of participating residents. Newsletters to staff were accompanied by nominal gift cards to express appreciation.

Figure 3. Sample page from the quarterly newsletter mailed to Ambient Independence Measures for Guiding Care Transitions (AIMS) study participants and staff. Study enrollment at the time of this newsletter was 89 residents and 20 staff.



ORCATECH
SENSING LIFE KINETICS

AIMS STUDY
QUARTERLY OVERVIEW

FEBRUARY 2016 – ISSUE 1

AIMS STUDY PARTICIPANTS AT A GLANCE

89 resident participants
- **80% are female**
- **Average age is 87 years**
- **Average education is 16 years**

These participants reside in one of seven CCRCs taking part in our study.

20 CCRC staff participants
- **90% female**
- **Average age is 41.1 years**
- **Average years at CCRC is 5.5**
- **Average years in their field is 11.4**

YOUR SLEEP HABITS

24% of you get up before 6:30am

17% of you go to bed after 11pm

The average AIMs participant sleeps for **9 hours and 2 minutes** each night.

WHAT IS THE AIMS STUDY?

The purpose of the Ambient Independence Measures for Guiding Care Transitions (AIMs) study is to determine how the information we gather from ORCATECH's in-home technology can provide retirement community staff more information about the health of their residents.

Sometimes it's not clear that someone could use additional help until they have a bad fall or other incident, and then it may be too late to avoid having to move to assisted living or nursing care. We hope our technology may eventually help residents receive assistance earlier so these situations can be avoided.

WHAT HAVE WE LEARNED SO FAR?

Since we are just finishing up enrolling participants, we've mostly learned about YOU!

How many online weekly forms have study participants completed?

- 96% of study participants completed at least 1 health form during December 2015 with an average of 3.24 forms per person for the month.
- Over 1400 online weekly health forms have been completed by our study participants between February and October 2015!

Look forward to more data updates in our next issue!

Data Collection and Analysis

For the duration of the study, dashboard access metrics were tracked, including the number of log-ins, average time spent on each dashboard page, and number of page views by each staff member.

Staff were emailed monthly surveys querying their use of the AIMS dashboard. They were asked whether they had looked at the dashboard in the last month and whether the dashboard was used in discussions about any residents. If they reported not having used the dashboard, they were asked to provide a reason. All surveys ended with an open request for feedback about the web-based activity dashboard. Midway through the study, in an effort to reengage, staff participants were sent a new survey with sample data illustrating acute and subtle changes in the behavior of one study resident. The staff responded to questions regarding the management of alerts and subsequent actions. For example, they were asked to identify preferred methods for the transmission of event reports, whether by dashboard alerts, emails, or other formats. Possible time frames for alerts, follow-up actions, and data interpretations were also probed.

At the end of the 36-month data collection period of the study, interviews with staff participants were held at 2 facilities jointly by 2 authors (KW and J Kaye) to discuss actual dashboard use among staff, study-related workload, and any other factors related to their use of the dashboard over the course of the study. The interviews were open ended and intended to elucidate barriers and opportunities for improving staff engagement in future studies (Multimedia Appendix 1).

Descriptive data for participating residents were collected at baseline. The rest of the data presented here were collected via web-based surveys and in-person interviews. Owing to small numbers, quantitative analyses of staff responses were not deemed appropriate.

Human Subjects' Protections

The protocol was approved by the Oregon Health & Science University Institutional Review Board (IRB#9944). Written informed consent was obtained from all participants before their inclusion in the study. The older adult residents did not receive compensation for study participation; the staff received nominal gift cards with newsletter mailings.

Results

Participants

A total of 95 older adult residents from 7 residential facilities in the Portland, Oregon, metro area were recruited and enrolled into the AIMS cohort (Table 2). They were 80% (76/95) female, with a mean age of 86.4 years (range 70-105 years) and Mini-Mental State Examination score of 28.7 (range 21-30). Overall, 10 participants had a Clinical Dementia Rating score of 0.5, consistent with mild cognitive impairment. Between enrollment and the data collection period, 4 participants withdrew; of the remaining 91 participants, 44 were assigned to the *viewable data* group and 47 to the *nonviewable data* group.

During data collection from December 2014 to December 2017, 6% (6/95) AIMS participants made a permanent move from independent living to assisted living or to a health care center. In addition, 32% (30/95) participants answered “yes” at least once to the weekly question regarding new care provision. The most commonly reported forms of assistance were medication management (n=22) and help with bathing, dressing, and grooming (n=19). Most of the additional assistance was provided by facility staff (n=20), whereas a minority received assistance from family or privately hired caregivers.

Table 2. Baseline demographic and clinical characteristics (N=95).

Variable	Value
Age (years), mean (SD)	86.4 (7.4)
Female, n (%)	76 (80)
Education (years), mean (SD)	15.9 (2.4)
Mini-Mental State Examination, mean (SD)	28.7 (1.6)
Cognitively impaired, n (%)	10 (11)
Geriatric Depression Scale, mean (SD)	1.1 (1.9)
Functional Assessment Questionnaire, mean (SD)	1.0 (3.3)
Cumulative Illness Rating Scale, mean (SD)	20.6 (2.6)

Staff Engagement

During the 3 years of study, 11 of the 25 consented staff members logged in at least once to the activity dashboard (Table 3). All facilities were represented by at least one staff member. The number of unique log-ins to the dashboard per facility ranged from 1 to 9. Staff *page views*, that is, the number of pages of data looked at per staff member, ranged from 4 to 211 over the duration of data collection. It is clear that some staff were more engaged than others; at facility #7, 1 of 5 participating staff members logged in during the course of the study, but that staff member had 211 page views. In the 4 facilities where more than one staff member viewed the dashboard, there was consistent overlap in the residents viewed. For example, in one facility, 2 staff members viewed the same 3 residents’ data. Some residents merited multiple page views. Of the 6 residents who actually transitioned to a higher level of care during the study period, 4 had been randomly assigned to the group where monitored data were shared with staff; only 2 of these transitioned residents’ data were viewed by the staff before their transitions.

Monthly surveys were discontinued because of a low response rate. Across the 5 monthly email surveys, a total of 81 invitations to respond were sent to the staff. A total of 25 completed surveys were returned. Of the 25 surveys, in 12 instances, staff members indicated that they had looked at the web-based AIMS dashboard in the past month. Survey responses of those who had not used the dashboard in the past month (n=13) indicated that they forgot to use the dashboard (7/13, 54%), they did not need to use it because of their role in the organization (4/13, 30%), or they were unable to get onto the dashboard system because they forgot their password (2/13, 15%). Of the 12 instances where the staff indicated they had looked at the web-based AIMS dashboard, none had used the information from the dashboard as a part of a discussion about a resident they viewed.

Of the 25 eligible staff participants, 5 responded to the midstudy survey, which included sample data designed to reengage staff participants, emailed in January 2017. They generally endorsed a preference for controlling the frequency of initial alerts and follow-up reminders. Although 4 of the 5 respondents judged alerts to acute changes to be useful, only 2 respondents felt the same for alerts regarding subtle changes or trends. In open text, they explained that subtle changes in behavior were not likely to affect the overall function or well-being and were not *acute enough* to warrant their involvement. One respondent elaborated:

> *There is a fine line between monitoring someone’s independent lives and knowing when to interfere for safety reasons...It is hard to know when to involve a care team without being too Orwellian. I would likely wait a month and then have a bit more data to take to a team meeting to assess the subtle changes collectively.*

At the end of the data collection period, 6 staff participants at 2 facilities were interviewed by 2 authors (KW and J Kaye). Interviews ranged from 1 to 1.5 hours. Feedback regarding barriers to the use of the AIMS dashboard and the data presented fell into 2 general categories.

Table 3. Staff engagement with the data dashboard.

Site	Number of participants consented	Dashboard use		
		Total number of participants logged	Total number of unique log-ins	Total page views
1	5	2	3	121
2	3	2	6	57
3	2	2	2	51
4	3	2	9	88
5	4	1	1	17
6	3	1	2	40
7	5	1	4	211

Practical Issues and Barriers

Technical difficulties with the use of dashboard at the beginning of the study proved to be dissuasive for some staff members:

> *We had difficulty logging into the system in the beginning.*

The infrequency of use exacerbated the log-in challenges because of continued unfamiliarity with the system. Furthermore, multiple staff noted that once they did access the site, the residents of interest to them in terms of changing care needs were not always study participants (because a minority of the residents they were overseeing were in the study at their site), making it less likely that they would re-enter the site. At the same time, the staff felt inundated by data in general and lacked the time to adequately review and interpret dashboard metrics. More than one staff commented that the study required a designated staff member to monitor data. Alternatively, one staff member suggested that if alerts were triaged to appropriate staff, they would all receive fewer irrelevant emails and alerts. Another staff member added that it would be critical for the system's effectiveness that staff only receive alerts relevant to their position, "then if you get an alert you know it was meant for you." The frequency of alerts and potential false alarms were naturally of concern in relation to time management for already overextended care providers.

Other feedback related to specific behaviors was monitored by the AIMS project. Despite initial enthusiasm about the areas of interest (eg, sleep, medication adherence, socialization), the staff subsequently recognized additional behaviors as more relevant to their decision making, such as disruptive behaviors or missing meals or appointments. At the same time, others appreciated receiving *real-time data* on metrics such as weight and sleep duration. They did acknowledge that although the residents were excited to be part of a research program, the staff felt they needed more experience with possible outcomes to see the benefits of behavior monitoring. In general, the staff struggled with the challenge of responding to acute events versus detecting trends and patterns of behavioral decline and determining how to integrate such monitoring into their daily schedules. Ultimately, some saw these data as potentially helpful in developing a model for transition, allowing them to be more proactive and less reactive.

Professional Bias and Ethical Concerns

One interviewee acknowledged that she had biases from the beginning, in that her training as a counselor led her to be more intuitive than data driven in her decision making. Others noted their strong inclination to use the data as objective support for their own subjective perceptions of care needs.

A second concern was related to the inherent conflict between resident autonomy and safety. Multiple staff members voiced this sentiment, citing the necessary compromise between letting a resident "do what they want" even if that were to include risky behavior. However, the installation of monitoring technology raised the issue of risk management for some, in that knowledge of potentially unsafe behavior would require a decision regarding the appropriate staff response. Patient autonomy and privacy were referenced in a question posed by a director of nursing services: "How paternalistic do you want your environment to be?" An intrinsic tension between residents' desire for control and their general willingness to share monitoring data was reflected in staff efforts to provide optimal care while respecting self-determination.

Discussion

Principal Findings

We report the results of developing and implementing an automated, continuous data monitoring platform that presented CCRC residents' daily activity data on a regular basis to professional staff charged with decisions about care transitions. Our goal was to assess whether these activity metrics meaningfully contributed to this decision-making process and to test their contribution by examining in a randomized controlled trial framework whether those metrics might inform decisions regarding transitions to higher levels of care by providing early and actionable data on changes in behavior and activity.

During the 3 years of study monitoring, only 6 participants transitioned to a higher level of residential care. This number was lower than anticipated and may reflect a growing trend toward engaging additional in-home assistance instead. A total of 30 participants reported needing new in-home assistance during the study, ranging from medication management to assistance with bathing. The low rate of transitions may have contributed to the underutilization of the monitoring data dashboard by facility care staff.

Although initial acceptance of the project was enthusiastic, with the staff in 7 facilities committing to regular utilization of the data dashboard, this enthusiasm failed to carry over into implementation. The staff used the dashboard sporadically, and those who did identified several limitations to use, ranging from technology challenges to ethical concerns.

Comparisons With Previous Work

Previous work on the adoption of technology by health care professionals has used various iterations of the Technology Acceptance Model (TAM) [23]. This model postulates that the intention to use technology is predicated on attitudes that are mediated by the perceived attributes of the technology. The two most important attributes in explaining acceptance and use of technology have been proposed to be perceived usefulness and perceived ease of use [24]. Modified technology acceptance models have added subjective norms and facilitator conditions as determinants of intent to use [25].

Although all models have been shown to have some explanatory power, research applications have typically incorporated *behavioral intention to use* rather than actual use. Few studies include measures of actual technology use, relying instead on measures of behavioral intent. Our experience in this project suggests that despite perceived usefulness at baseline and attempts to accommodate users' needs to achieve perceived ease of use, actual technology use can still be lower than predicted by indicators of intent to use.

TAMs have been applied to the identification of barriers to adoption by health care providers. Organizational factors such as administrative leadership and support, including additional time allotment and clear incentives, adequate resources for training and ongoing technical support, and organizational planning for implementation have been cited as important barriers [26-29]. Technical impediments include malfunctioning or unreliable equipment and devices and lack of coordination or complementarity with existing procedures. Failure to include potential end users in the design and planning of technology applications has also been cited as an impediment to adoption [27,30]. In addition, the ability of users to exert control over the technology's behavior has been cited as an important motivator in the adoption of a new technology [31]. Sabrowski and Kollak [32] describe the *domestication* of technology as a process whereby the system or device is integrated and adapted to the user's needs and environment. They postulate that until care professionals view a technology as integral to an improvement in the delivery of care, they will be resistant to adoption. Finally, human factors connected to attitudes and previous experiences with technology can have enormous influence. Lack of knowledge or familiarity with a device or system can diminish both perceived usefulness and ease of use. Furthermore, for health care providers seeking to maximize the quality of life for a medically fragile population, concerns about loss of human contact can foster negative attitudes toward technology. Savenstedt et al [33] identified themes elicited from interviews with professional caregivers of older adults. Technology applications were seen as both an aid and a threat to not only humane care but also to their roles as caregivers. They cited the loss of immediate contact and involvement with their care recipients as a potential consequence of technology applications. The authors suggest that these inherent conflicts foster resistance to change despite outward acceptance.

Limitations and Lessons Learned

We found sporadic adoption of a new monitoring technology by professional staff. Despite initial enthusiasm and ongoing efforts to engage the participating staff in 7 residential care communities, the goal of this study, that is, to analyze the impact of technology-based data on decision making around transitions in care, was not achieved. Previous research has described organizational, personal, and technological characteristics and contexts that may facilitate or impede the adoption of health technologies [29,34-37]. Feedback from our staff participants was consistent with these barriers to technology use.

Organizational Barriers to Use

Organizational factors such as clearly communicated expectations and possible study outcomes, continuous monitoring of technical support needs of the staff, and recognition of time commitments may have been inadequately addressed. A consistent recommendation has been the early inclusion of end users in design and implementation. Although our initial focus groups elicited some preferences and priorities, a longer run-in iterative process might have reduced the gap between our efforts at participatory design and the reality of the final implementation. Although some staff members recommend identifying a *champion* or *super-user* at each facility to provide onsite, continuous support and motivation to engage with the platform [37], we found few staff members who self-identified as such. Other organizational factors, although beyond the scope of this project, should be considered moving forward. A unified commitment to the implementation of new technologies by the administration, staff, and residents is required and must include the provision of adequate time for staff education and training, recognition of professional autonomy, and ongoing identification of potential barriers.

Personal and Professional Barriers to Use

Personal and professional traits related to technology adoption include experience with technology, peer attitudes, staff engagement, and professional satisfaction [29,34]. Although initial training was provided to all staff in dashboard use and features, additional active ongoing technical support might have increased engagement. Perceived usefulness may have been diminished because of lack of concordance between resident research participants and particular residents of interest (who were not monitored in the study) to the transition teams.

The staff expressed ethical concerns related to the quality of care and privacy. Previously, unexamined conflicted attitudes toward monitoring technology surfaced only after actual engagement with the system. A more detailed discussion of the implications and possible outcomes of staff participation in a continuous monitoring study before implementation might have mitigated ethical apprehensions. An increase in staff workload or change in procedure, without adequate motivation and explanation, cannot be expected to be enthusiastically adopted.

In the future, identification of intrinsic and extrinsic motivators, adequate and sustained training, and a realistic understanding of the goals of the study must be an integral part of research in technology adoption.

Technical Barriers to Use

Finally, issues with the technology itself may have presented barriers to implementation. Technical issues such as failed passwords and initial platform malfunctions, while infrequent, led to some early negative interactions, which proved to be difficult to overcome. In addition to the initial discussion of needs with end users, further refinement of the platform might have enhanced sustained participation. However, the refinement of protocols must be balanced with the time and resources needed to implement a program, acknowledging that staff turnover can attenuate involvement over time.

The original premise of our intervention was that a less obtrusive, information-on-demand approach would be least disruptive to workflows. However, at least initially, sending notifications of changes in resident behavior rather than relying on the staff to engage and retrieve data from a novel and unfamiliar source might have increased their understanding of the utility of the technology and its relevance to their daily practice. Providing actionable, customized information on residents at risk would demonstrate the potential benefits of continuous monitoring over standard procedures regarding transitions in care.

Conclusions

The limitations of previous work describing the intent to use technology without the inclusion of actual use as the final outcome are demonstrated by our findings. Initial enthusiasm and support for in-home, continuous monitoring of activity and behavior was established among the staff of 7 continuing care residential communities. Nevertheless, multiple factors, whether technical, personal, or institutional, intervened between intent and use. Future research examining technology adoption cannot ignore this crucial outcome measure if widespread acceptance and implementation of health care technologies are to be advanced. Finally, and perhaps most importantly, future work should examine whether a culture change toward proactive intervention to prevent or safely delay unwanted care transitions, rather than using technology for emergency response and acute situational management, will achieve wider use of technologies across residential care communities and related settings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
End of study discussion guide.
[DOCX File , 13 KB - jmir_v23i1e18806_app1.docx]

References

1. Callahan CM, Tu W, Unroe KT, LaMantia MA, Stump TE, Clark DO. Transitions in care in a nationally representative sample of older Americans with dementia. J Am Geriatr Soc 2015 Aug;63(8):1495-1502 [FREE Full text] [doi: 10.1111/jgs.13540] [Medline: 26200764]
2. Fong JH, Mitchell OS, Koh BS. Disaggregating activities of daily living limitations for predicting nursing home admission. Health Serv Res 2015 Apr;50(2):560-578 [FREE Full text] [doi: 10.1111/1475-6773.12235] [Medline: 25256014]
3. Miller LM, Dieckmann NF, Mattek NC, Lyons KS, Kaye JA. Social activity decreases risk of placement in a long-term care facility for a prospective sample of community-dwelling older adults. Res Gerontol Nurs 2014;7(3):106-112 [FREE Full text] [doi: 10.3928/19404921-20140110-02] [Medline: 24444452]
4. Nielsen AB, Siersma V, Waldemar G, Waldorff FB. The predictive value of self-rated health in the presence of subjective memory complaints on permanent nursing home placement in elderly primary care patients over 4-year follow-up. Age Ageing 2014 Jan;43(1):50-57. [doi: 10.1093/ageing/aft131] [Medline: 23985335]
5. Kelsey SG, Laditka SB, Laditka JN. Dementia and transitioning from assisted living to memory care units: perspectives of administrators in three facility types. Gerontologist 2010 Apr;50(2):192-203. [doi: 10.1093/geront/gnp115] [Medline: 19633203]
6. Lieto JM, Schmidt KS. Reduced ability to self-administer medication is associated with assisted living placement in a continuing care retirement community. J Am Med Dir Assoc 2005;6(4):246-249. [doi: 10.1016/j.jamda.2005.04.002] [Medline: 16005410]
7. Maxwell CJ, Soo A, Hogan DB, Wodchis WP, Gilbart E, Amuah J, et al. Predictors of nursing home placement from assisted living settings in Canada. Can J Aging 2013 Dec;32(4):333-348. [doi: 10.1017/S0714980813000469] [Medline: 24180746]
8. Rosenberg PB, Mielke MM, Samus QM, Rosenblatt A, Baker A, Brandt J, et al. Transition to nursing home from assisted living is not associated with dementia or dementia-related problem behaviors. J Am Med Dir Assoc 2006 Feb;7(2):73-78. [doi: 10.1016/j.jamda.2005.06.003] [Medline: 16461247]

9. Couture M, Ducharme F, Lamontagne J. The role of health care professionals in the decision-making process of family caregivers regarding placement of a cognitively impaired elderly relative. Home Health Care Manag Pract 2012 Apr 11;24(6):283-291. [doi: 10.1177/1084822312442675]
10. Georgiou A, Marks A, Braithwaite J, Westbrook JI. Gaps, disconnections, and discontinuities-the role of information exchange in the delivery of quality long-term care. Gerontologist 2013 Oct;53(5):770-779. [doi: 10.1093/geront/gns127] [Medline: 23103520]
11. Demiris G, Thompson H. Smart homes and ambient assisted living applications: from data to knowledge-empowering or overwhelming older adults? Contribution of the IMIA smart homes and ambiant assisted living working group. Yearb Med Inform 2011;6:51-57. [Medline: 21938325]
12. Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent systems for assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66(Suppl 1):i180-i190 [FREE Full text] [doi: 10.1093/geronb/gbq095] [Medline: 21743050]
13. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive computing technologies to continuously assess alzheimer's disease progression and intervention efficacy. Front Aging Neurosci 2015;7:102 [FREE Full text] [doi: 10.3389/fnagi.2015.00102] [Medline: 26113819]
14. Austin J, Klein K, Mattek N, Kaye J. Variability in medication taking is associated with cognitive performance in nondemented older adults. Alzheimers Dement (Amst) 2017;6:210-213 [FREE Full text] [doi: 10.1016/j.dadm.2017.02.003] [Medline: 28349120]
15. Hayes TL, Larimer N, Adami A, Kaye JA. Medication adherence in healthy elders: small cognitive changes make a big difference. J Aging Health 2009 Jun;21(4):567-580 [FREE Full text] [doi: 10.1177/0898264309332836] [Medline: 19339680]
16. Dodge HH, Mattek NC, Austin D, Hayes TL, Kaye JA. In-home walking speeds and variability trajectories associated with mild cognitive impairment. Neurology 2012 Jun 12;78(24):1946-1952 [FREE Full text] [doi: 10.1212/WNL.0b013e318259e1de] [Medline: 22689734]
17. Akl A, Chikhaoui B, Mattek N, Kaye J, Austin D, Mihailidis A. Clustering home activity distributions for automatic detection of mild cognitive impairment in older adults. J Ambient Intell Smart Environ 2016;8(4):437-451 [FREE Full text] [doi: 10.3233/AIS-160385] [Medline: 27617044]
18. Mancini M, Schlueter H, El-Gohary M, Mattek N, Duncan C, Kaye J, et al. Continuous monitoring of turning mobility and its association to falls and cognitive function: a pilot study. J Gerontol A Biol Sci Med Sci 2016 Aug;71(8):1102-1108 [FREE Full text] [doi: 10.1093/gerona/glw019] [Medline: 26916339]
19. Hayes TL, Riley T, Mattek N, Pavel M, Kaye JA. Sleep habits in mild cognitive impairment. Alzheimer Dis Assoc Disord 2014;28(2):145-150 [FREE Full text] [doi: 10.1097/WAD.0000000000000010] [Medline: 24145694]
20. Kaye J, Mattek N, Dodge HH, Campbell I, Hayes T, Austin D, et al. Unobtrusive measurement of daily computer use to detect mild cognitive impairment. Alzheimers Dement 2014 Jan;10(1):10-17 [FREE Full text] [doi: 10.1016/j.jalz.2013.01.011] [Medline: 23688576]
21. Seelye A, Mattek N, Howieson DB, Austin D, Wild K, Dodge HH, et al. Embedded online questionnaire measures are sensitive to identifying mild cognitive impairment. Alzheimer Dis Assoc Disord 2016;30(2):152-159 [FREE Full text] [doi: 10.1097/WAD.0000000000000100] [Medline: 26191967]
22. Seelye A, Mattek N, Sharma N, Witter P, Brenner A, Wild K, et al. Passive assessment of routine driving with unobtrusive sensors: a new approach for identifying and monitoring functional level in normal aging and mild cognitive impairment. J Alzheimers Dis 2017;59(4):1427-1437 [FREE Full text] [doi: 10.3233/JAD-170116] [Medline: 28731434]
23. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
24. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
25. Venkatesh V, Morris M, Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
26. de Veer AJ, Francke AL. Attitudes of nursing staff towards electronic patient records: a questionnaire survey. Int J Nurs Stud 2010 Jul;47(7):846-854. [doi: 10.1016/j.ijnurstu.2009.11.016] [Medline: 20022007]
27. Weber S, Crago EA, Sherwood PR, Smith T. Practitioner approaches to the integration of clinical decision support system technology in critical care. J Nurs Adm 2009 Nov;39(11):465-469. [doi: 10.1097/NNA.0b013e3181bd5fc2] [Medline: 19898097]
28. Whitten P, Holtz B, Meyer E, Nazione S. Telehospice: reasons for slow adoption in home hospice care. J Telemed Telecare 2009;15(4):187-190. [doi: 10.1258/jtt.2009.080911] [Medline: 19471030]
29. Tak SH, Benefield LE, Mahoney DF. Technology for long-term care. Res Gerontol Nurs 2010 Jan;3(1):61-72 [FREE Full text] [doi: 10.3928/19404921-20091103-01] [Medline: 20128544]
30. Wild K, Boise L. In-home monitoring technologies: Perspectives and priorities of older adults. In: Augusto JC, editor. Handbook of Ambient Assisted Living. Amsterdam: IOS Press; 2012.
31. Wu P. A mixed methods approach to technology acceptance research. J Assoc Inf Syst 2012 Mar;13(3):172-187. [doi: 10.17705/1jais.00287]

32. Saborowski M, Kollak I. 'How do you care for technology?' – Care professionals' experiences with assistive technology in care of the elderly. Technol Forecast Soc Change 2015 Apr;93:133-140. [doi: 10.1016/j.techfore.2014.05.006]
33. Sävenstedt S, Sandman PO, Zingmark K. The duality in using information and communication technology in elder care. J Adv Nurs 2006 Oct;56(1):17-25. [doi: 10.1111/j.1365-2648.2006.03975.x] [Medline: 16972914]
34. Doran D, Haynes BR, Estabrooks CA, Kushniruk A, Dubrowski A, Bajnok I, et al. The role of organizational context and individual nurse characteristics in explaining variation in use of information technologies in evidence based practice. Implement Sci 2012 Dec 31;7:122 [FREE Full text] [doi: 10.1186/1748-5908-7-122] [Medline: 23276201]
35. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-e13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
36. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
37. Schoville RR. Discovery of implementation factors that lead to technology adoption in long-term care. J Gerontol Nurs 2017 Oct 1;43(10):21-26. [doi: 10.3928/00989134-20170914-06] [Medline: 28945269]

Abbreviations

AIMS: Ambient Independence Measures for Guiding Care Transitions
CCRC: continuing care retirement community
TAM: Technology Acceptance Model
ORCATECH: Oregon Center for Aging & Technology

Edited by G Eysenbach; submitted 19.03.20; peer-reviewed by N Thomas, A Piau, N Shen; comments to author 12.06.20; revised version received 23.07.20; accepted 15.09.20; published 13.01.21.

Please cite as:
Wild K, Sharma N, Mattek N, Karlawish J, Riley T, Kaye J
Application of In-Home Monitoring Data to Transition Decisions in Continuing Care Retirement Communities: Usability Study
J Med Internet Res 2021;23(1):e18806
URL: https://www.jmir.org/2021/1/e18806
doi:10.2196/18806
PMID:33439144

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org